MOUNT SINAI
GUSTAVE L.
AND
JANET W.
LEVY
LIBRARY

MOUNT SINAI
GUSTAVE L.
AND
JANET W.
LEVY
LIBRARY

THE MOUNT SINAI JOURNAL OF MEDICINE

OLUME 63 NUMBER 1 JANUARY 1996

APR 24

CONTENTS

GRAND ROUNDS

Benign Prostatic Hyperplasia: Natural History, Endocrine Pathogenesis, and Medical Treatment with Finasteride
Louis M. Sherwood 1

Thyroid Nodules and the Detection of Thyroid Cancer
Richard S. Haber 10

The Amyloidoses *Joel Buxbaum* 16

ART AND SCIENCE OF CONSULTATION

Consultation in Sleep Medicine *Lee K. Brown* 24

Evolution of Cardiologic Consultation *Leslie A. Kuhn* 39

GENERAL ARTICLES

Involuntary Confinement for Tuberculosis Control: The Jewish View *Fred Rosner* 44

The Need for More Primary Care Physicians to Contain Costs and Improve Access: Myths Versus Realities *Barry Stimmel* 49

Editor's Comment *Sherman Kupfer, Editor-in-Chief* 60

An Imaginary Conversation with the Gang of Three: A Ghostly Interview with Burrill B. Crohn, Leon Ginzburg, and Gordon Oppenheimer *Henry D. Janowitz* 61

BOOK REVIEWS

How to Quit Smoking without Gaining Weight
E. Neil Schachter 66

continued inside

STATEN ISLAND UNIVERSITY HOSPITAL
Affiliated with The Mount Sinai Hospital and Mount Sinai School of Medicine

Veterans Affairs

THE MOUNT SINAI JOURNAL OF MEDICINE

The Mount Sinai Journal of Medicine is published by The Mount Sinai Medical Center of New York and has the following affiliates: Bronx Veterans Affairs Medical Center, New York; Elmhurst Hospital Center, Queens, NY; Queens Hospital Center, Queens, NY; and Staten Island University Hospital, Staten Island, NY.

Editor-in-Chief
Sherman Kupfer, M.D.

Editor Emeritus
Lester R. Tuchman, M.D.

Associate Editors
Harriet S. Gilbert, M.D.   Julius Wolf, M.D.

Administrative Assistant
Joseline Ilagan

Assistant Editor for Housestaff
Matthew J. Kapklein, M.P.H., M.D.

Assistant Editors
David H. Bechhofer, Ph.D.
Constanin A. Bona, M.D., Ph.D.
Edward J. Bottone, Ph.D.
Lewis Burrows, M.D.
Christopher P. Cardozo, M.D.
Mark R. Chassin, M.D., MPH, MPP
Joseph S. Eisenman, Ph.D.
Richard A. Frieden, M.D.
Frederick Friedman, M.D.
Steven Fruchtman, M.D.
Paul L. Gilbert, M.D.
James H. Godbold, Ph.D.
Joseph J. Guarneri, Ph.D.
Richard S. Haber, M.D.
Noam Harpaz, M.D.
Tomas Heimann, M.D.
Sheldon Jacobson, M.D., FACEP
Barry W. Jaffin, M.D.
Andrew S. Kaplan, D.D.S.
Samuel Kenan, M.D.
Suzanne Carter Kramer, M.Sc.
Mark G. Lebwohl, M.D.
Zvi Lefkowitz, M.D.
Kenneth V. Lieberman, M.D.
Lynda R. Mandell, M.D., Ph.D.
Steven Markowitz, M.D.
Bernard Mehl, D.P.S.
Leonard E. Meiselas, M.D.
Myron Miller, M.D.
Edward Raab, M.D.
Elliot J. Rayfield, M.D.
Allan Reed, M.D.
Clive Rosendorff, M.D., Ph.D.
Fred Rosner, M.D.
Allan E. Rubenstein, M.D.
David B. Sachar, M.D.
Henry Sacks, M.D.
Ira Sanders, M.D.
Martin H. Savitz, M.D.
Clyde B. Schechter, M.D.
Stuart C. Sealfon, M.D.
Michael Serby, M.D.
Phyllis Shaw, Ph.D.
Alex Stagnaro-Green, M.D.
Barry D. Stimmel, M.D.
Nelson Stone, M.D.
Max Sung, M.D.
Rein Tideiksaar, Ph.D.
Richard P. Wedeen, M.D.

Editorial Board
Barry S. Coller, M.D.
Barry Freedman, M.B.A.
Richard Gorlin, M.D.
Nathan Kase, M.D.
Panayotis G. Katsoyannis, Ph.D.
Charles K. McSherry, M.D.
Jack G. Rabinowitz, M.D.
John W. Rowe, M.D.
Alan L. Schiller, M.D.
Alan L. Silver, M.D.
Alvin S. Teirstein, M.D.
Rosalyn S. Yalow, Ph.D.

The Mount Sinai Journal of Medicine (ISSN No. 0027-2507; USPS 284-860) is published 6 times a year in January, March, May, September, October, and November in one indexed volume by The Mount Sinai Medical Center, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Subscription price: individuals (U.S. and Canada), $75 per year; libraries and institutions (U.S. and Canada), $115 per year; individuals (outside U.S. and Canada), $100 per year; libraries and institutions (outside U.S. and Canada), $140 per year. Single copies $20. New subscriptions begin with the first issue of the calendar year. Send notice of change of address 60 days before its effective date. Second-class postage paid at New York, NY and at additional mailing offices. POSTMASTER: Send address changes to The Mount Sinai Journal of Medicine, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Telephone number (212) 241-6108; Fax (212) 722-6386. Copyright 1996 The Mount Sinai Hospital.

THE
MOUNT SINAI JOURNAL
OF MEDICINE

Volume 63
Number 1
January 1996

CONTENTS *continued*

The PDR Family Guide to Women's Health and Prescription Drugs *Erica Asch* 66

Eccentric and Bizarre Behaviors *James J. Strain* 67

The Living Heart Brand Name Shopper's Guide *Lori Silver* 68

Analyzing Survival Data from Clinical Trials and Observational Studies *Sylvan Wallenstein* 68

Index to Advertisers

Southampton Hospital and Imugen
Conference on Lyme Disease inside back cover

Digitized by the Internet Archive
in 2015

https://archive.org/details/mountsinaijourna6319moun

Grand Rounds

Benign Prostatic Hyperplasia:

Natural History, Endocrine Pathogenesis, and Medical Treatment with Finasteride

Louis M. Sherwood, M.D.

This brief review represents an internist's view of benign prostatic hyperplasia (BPH), its natural history and epidemiology, the evidence that suggests it has an endocrine basis, and recent data concerning its medical management with finasteride.

Natural History and Epidemiology

What are the risk factors for BPH? Although other risk factors have been suggested, only increasing age and normal male androgenic function have been clearly established so far (1). Because most older men of various ethnic groups have these risk factors, a significant number of them are obviously at risk for this common disorder. The natural history and epidemiology of BPH are not nearly as well documented as those of other diseases such as atherosclerosis. There are limited longitudinal as well as cross-sectional epidemiologic data on BPH, but new information is being regularly obtained. There is much to be gained from further investigation of the natural history and epidemiology of BPH, because that is the only way to measure the value and outcomes of different interventions for BPH, whether they be medical, surgical, or watchful waiting. As physicians, we are a key part of the ongoing evolution of health care in the United States. Thus, one of our responsibilities is to consider the effectiveness of different therapies and to interpret the results of intervention in the context of the natural history of the disease.

With respect to anatomic BPH, Berry et al. (2) summarized several autopsy series on prostate size and histologic incidence of BPH. They reported that the prostate grows rapidly during adolescence, reaches adult size in the third decade, and then grows slowly until men are in their 40s and 50s when it begins to increase progressively in size (**Fig. 1**). They also determined that histologic BPH increases in incidence with age, starting in the third decade of life, and becomes universal by the time a man is in his 80s or 90s. If one examines prostatectomy specimens from men with BPH, one observes a heterogeneous admixture of epithelial and stromal hyperplasia (3).

With respect to the incidence of clinical BPH, important longitudinal data have been obtained

---

Adapted from a Grand Rounds presentation at The Mount Sinai Medical Center on November 2, 1993. Address reprint requests to L.M. Sherwood, M.D., Senior Vice President, Medical and Scientific Affairs, U.S. Human Health, Merck & Co., Inc., PO Box 4, Sumneytown Pike, West Point, PA 19486-0004.

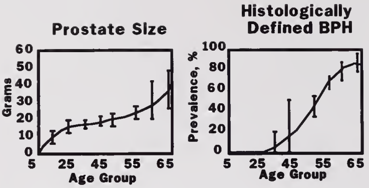


Fig. 1. Size of prostate and histologic presence of BPH with age from autopsy specimens. (Adapted from Berry et al. (1).)

from the extensive database of the Baltimore Longitudinal Study of Aging, in which approximately 1100 men in Baltimore were followed for up to 30 years (4). In the associated health questionnaire, questions pertaining to a weak stream and the sensation of incomplete emptying of the bladder were found to correlate with future prostatectomy in men with symptomatic BPH (see **Fig. 2**). In addition, these two symptoms progressed with age in most patients. Serum samples that were saved from this study showed that the rate of change in serum prostate specific antigen (PSA) began to increase in patients who were eventually diagnosed with prostate cancer, well before the cancer was clinically detectable.

Chute et al. (5) recently performed cross-sectional studies of more than 2000 men at random in Olmsted County, Minnesota, a fairly stable population that has been the subject of a number of epidemiologic studies. To establish the age-specific prevalence of urinary symptoms in this population, a questionnaire that was validated for assessing the severity of symptomatic BPH was administered to more than 2000 men in the community. **Figure 3** shows the percentage of men in each decade who had urinary symptoms that were compatible with moderate to severe BPH. These symptoms were not only quite common, but they also increased in frequency with each decade of life. Validated urologic questionnaires are most useful, of course, in quantifying the level of symptoms in men with diagnosed BPH and have been applied widely for that purpose.

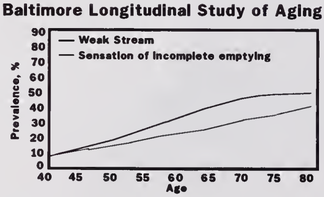


Fig. 2. Increase of BPH symptoms with age in Baltimore Longitudinal Study of Aging. (Adapted from Arrighi et al. (3).)

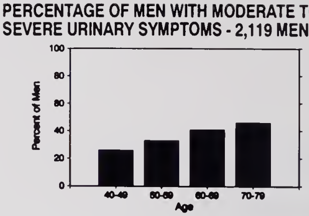


Fig. 3. Percentage of men in Rochester, Minnesota, community study having symptoms compatible with moderate to severe BPH. (Adapted from Chute et al. (5).)

Urinary flow rate can be measured by a relatively simple technique in which men void into a uroflowmeter, which is essentially a funnel connected to a flowmeter. Maximum urinary flow rate reflects the physiologic effects of both the resistance of the prostate and bladder neck to urinary flow, as well as the bladder's contractility. In the Olmsted County studies of urinary symptoms in men, Girman et al. (6) showed that the maximum urinary flow rate for men in their 40s is significantly higher than it is for men in their 70s (5). Maximum urinary flow rate declined slowly but progressively from about 20 mL/s to less then 15 mL for men in their 70s. The so-called obstructed range for urinary flow rate is below 10 mL/s, but overall, flow rate decreases slowly at about 2 to 3 mL/decade. Thus, over several years, a clinically meaningful decline in this parameter could occur. This is similar to other chronic diseases, such as hypertension, hyperlipidemia, or diabetes, where progression of certain complications is slow and insidious over a number of years.

In discussing the natural history of BPH, it is important to note that the progression of urinary symptoms is an important part of the pathophysiology of this disease that has not been emphasized sufficiently. This does not mean that symptoms progress in every man, but if you look at population data, it is clear that symptomatic BPH is a progressive disorder in many men.

Endocrine Pathogenesis of BPH

There is considerable evidence that BPH has an endocrine basis. Although prepubertal castration is no longer performed, it is known to have occurred in ancient societies, and the boys who had been castrated before puberty did not develop BPH as adults. Likewise, genetic diseases that impair androgen action or production are associated with inhibition of prostate growth (7, 8). Androgen ablation, by medical therapy or surgical castration, leads to shrinkage of the hyperplastic prostate (9, 10). Thus, there is considerable evidence from endocrine manipulations (usually involving medical or surgical castration), that there is a close relationship of androgens to BPH. The principal prostatic androgen is dihydrotestosterone (DHT), a reduced metabolite of testosterone. Initially it was felt that the DHT content of BPH tissue might be higher than in normal prostates. Although it is not clear that this is the case, DHT levels within the prostate remain at a normal level with aging, despite a slow decrease in plasma testosterone (11).

There is much more DHT than testosterone in the prostate. DHT is also the primary male androgen present in the nucleus of the prostate cell, and it is a more potent androgen than testosterone, because it has a greater binding affinity for the androgen receptor and a longer half-life (11). In tissues rich in 5-alpha-reductase, like the human prostate, high levels of DHT also result from a decrease in 3-hydroxysteroid reductase activity that metabolizes DHT. Thus, the human prostate is a favorable environment for the production and action of significant quantities of DHT.

The studies by Imperato-McGinley et al. (7) and Wilson et al. (8), furthered our understanding of the physiologic roles of DHT and testosterone, in the fetus and in the adult male. They independently described genetic male children, who were born with a primitive phallus, undescended testes, a blind vaginal pouch, and a rudimentary prostate, but otherwise had normal internal organs. Although genetic males, phenotypically they were pseudohermaphrodites, and due to the appearance of their genitalia at birth, they were often considered females at birth and were raised as such. Biochemically, these individuals had low levels of DHT relative to their testosterone levels, and it was postulated that they had a genetic abnormality manifested by decreased 5-alpha-reductase enzyme activity in the conversion of testosterone to DHT.

At puberty, their muscles hypertrophied, voices deepened, female-appearing external genitalia masculinized, a primitive phallus enlarged, testicular size increased, and they usually developed a male psychosexual orientation. What is especially noteworthy is that they did not develop acne or male pattern baldness in adult life and their prostates remained small, without BPH. In addition, it appeared that they did not have any other major metabolic problems (12).

It was thus determined, based on these and other reports, that testosterone in the male fetus is critical for differentiation of the Wolffian duct into the internal male sexual structures (epididymis, vas deferens, and seminal vesicles). During puberty and adulthood, testosterone seems to be the critical factor in the development of the male voice, laryngeal enlargement, male sexual function, libido, and sperm production. During puberty, testosterone is also required for further enlargement of the male sexual organs. DHT, on the other hand, seems to be essential for the normal fetal differentiation of the external "male" genitalia, and the development of the prostate. A male fetus deficient in dihydrotestosterone will therefore not have normal differentiation of the external genitalia and will be a pseudohermaphrodite at birth. In addition, males who are born deficient in DHT have a relatively rudimentary prostate that does not become enlarged with age.

What is the role of DHT after puberty and in adult life? Its physiologic role is not known, but we do know that normal circulating levels of DHT are associated with three syndromes: acne, male pattern baldness, and benign prostatic hyperplasia. In tissues rich in 5-alpha-reductase such as the prostate, levels of DHT are much higher than testosterone. Thigpen et al. (13) recently showed that there are actually two 5-alpha-reductase isoenzymes; type-2 enzyme predominates in the prostate whereas type-1 enzyme seems to be more prominent in the skin and in the liver. The individuals described by Imperato-McGinley et al. (7) and Wilson et al. (8) have a defect in the type-2 isoenzyme, and there is a variety of gene mutations leading to differences in phenotypic expression.

Although the data suggest that DHT is the major driver of prostate growth, there may be other factors that contribute as well. Estrogen and various growth factors are likely to play a supportive role in the etiology of BPH (11), but it is difficult to neglect data that strongly suggest that DHT is the key factor in prostate growth. To that end, a great deal of scientific work at the Merck Research Laboratories went into the discovery and development of a compound that would inhibit the 5-alpha-reductase enzyme that catalyzes the conversion of testosterone to DHT.

Clinical Development of Finasteride

Efficacy

The key factors that led Merck scientists to initiate a series of investigations to discover and develop an inhibitor of the 5-alpha-reductase enzyme were the genetic disorder 5-alpha-reductase deficiency and the evidence suggesting that BPH had an endocrine etiology (14). The hypothesis was that interference with the synthesis of DHT by a 5-alpha-reductase inhibitor might cause regression of the enlarged prostate gland and inhibit enlargement of the prostate, thereby blocking the progression of BPH. It was also hypothesized that a 5-alpha-reductase inhibitor would be able to control the hyperplasia of the prostate (because its growth was dependent on DHT) without significantly affecting sexual function, which is dependent on testosterone. This is in contrast to other hormonal manipulations to treat BPH, such as surgical castration or its medical equivalent using a gonadatrophin-releasing hormone agonist. Both are less desirable because they decrease both testosterone and DHT (9, 10).

The genetic disorder, 5-alpha-reductase deficiency, was reported in the 1970s when Merck scientists started working on the discovery of enzyme inhibitors. At that time it was not appreciated that there were two isoenzymes. In 1992, the drug finasteride was marketed for the treatment of symptomatic BPH. Why did it take so long to solve what seemed to be a relatively straightforward problem? The enzyme to inhibit the substrate was known, and it was thought that a steroidlike molecule would be a successful inhibitor. Several years passed during which many compounds were tested in the laboratory. Many of the early enzyme inhibitors were toxic, particularly to the liver of experimental animals. The compound that was developed, finasteride, is an azasteroid molecule (14) (see Fig. 4). It is closely related to steroid hormones in structure and is a time-dependent inhibitor of the 5-alpha-reductase enzyme. The compound inhibits both isoenzymes but is more potent in inhibiting the prostatic enzyme. Finasteride does not bind the androgen receptor and therefore does not have anti-androgenic properties.

The objectives of the large-scale pivotal clinical studies that were conducted prior to marketing were as follows (14): (1) to decrease circulating DHT levels; (2) to reduce prostatic size; (3) to improve urinary flow rate; (4) to improve symptoms; and (5) to maintain these changes over time. These clinical trials included two similar multicenter studies (North American and international) in which 1645 men were randomized into one of three groups: placebo, finasteride 1 mg, and finasteride 5 mg (all patients received placebo for 2 weeks during the placebo run-in period) (14). When these clinical studies were initiated, the optimal dose of finasteride was not known. When the studies were completed, finasteride 5 mg was shown to be more effective and to have the same safety profile as finasteride 1 mg. When these studies were initiated in 1988, they were the largest randomized, prospective, controlled clinical studies of BPH ever performed (15). At the end of the 1-year placebo-controlled period, men were offered an opportunity to continue on finasteride 5 mg. Over two thirds of the men elected either to continue to take finasteride 5 mg, or to switch from placebo or 1 mg to the 5-mg dose. The 36-month efficacy data for the open extension part of this study are from the cohort of patients who received finasteride 5 mg in the 12-month placebo-controlled study (16).

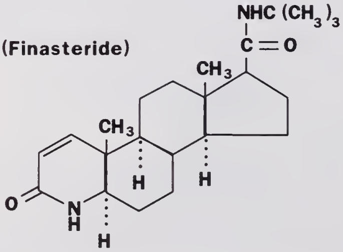


Fig. 4. Organic structure of finasteride. (From Stoner (14).)

The serum level of DHT decreases rapidly within days of administering finasteride, and it stays down as long as the drug is taken (15) (see Fig. 5). Serum DHT remained suppressed for men who were on finasteride for 36 months, at a median level of approximately 70% below baseline values (16). In one of the early clinical studies with finasteride, we determined the suppression of DHT by finasteride in prostate tissue (14). Men who were already scheduled for a prostatectomy for BPH were treated with finasteride for approximately 7 to 10 days, and then the DHT levels were assayed in the surgically removed prostate specimens. Prostatic DHT levels substantially decreased, whereas the prostatic testosterone levels

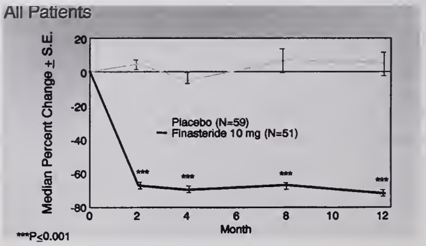


Fig. 5. Effect of finasteride compared with placebo on serum DHT levels: percent change from baseline. (Adapted from Gormley et al. (15).)

increased up to 10 times over pretreatment levels relative to placebo. The net androgen load in the prostate was decreased, however, because DHT is 5 to 10 times more potent than testosterone in binding to the androgen receptor.

To measure prostate volume in the North American trials, magnetic resonance imaging (MRI) was used because it was the most precise technique available. Internationally, transrectal ultrasound was used. Both methods are relatively accurate and comparable in determining relative prostatic volumes. In both studies, there was a median shrinkage of the prostate by approximately 20% at 1 year (15). The shrinkage was maintained in the open extension studies (16).

In a young man without BPH, the transition zone and its associated periurethral region comprise approximately 5% of the prostatic volume (1). Because BPH arises from the periurethral/transition zone, this area can become quite large in a man with BPH and, in fact, can dominate the prostate. Tempany et al. (17) recently published data from a 12-month placebo-controlled study that investigated the effect of finasteride 5 mg on the volume of the periurethral/transition zone, using MRI to measure prostatic volume. They showed that the prostatic regression due to the drug was essentially in the transition zone where BPH occurs. In another study, investigators are performing prostate biopsies on men before and after treatment with finasteride to assess the histologic changes that occur in the prostate after administration of the drug.

In the open extension studies, regression of the enlarged prostate was maintained, with the median regression being 27% at 36 months (16) (see **Fig. 6**). Although most prostatic regression occurred in the first year, some additional decrease in prostate volume may occur with continued therapy. The results of a long-term placebo-controlled study (3000 men for 4 years) now under way are necessary to validate the magnitude of any changes after 1 year.

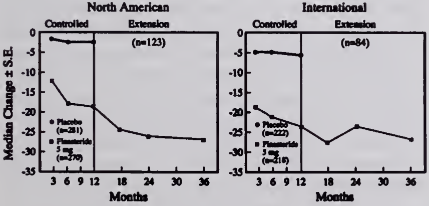


Fig. 6. Effect of finasteride on prostate volume over 36 months of treatment: median percent change from baseline. (From Stoner et al. (16).)

The next parameter examined was urinary flow rate (see **Fig. 7**). Because patients in the study periodically void into a urinary flow machine to measure urinary flow rates, there was probably some training effect that occurred, which may account (in part) for the placebo effect. The maximum flow rate curve of the placebo group reached an early plateau, however, with no further increase during the first year. Patients treated with finasteride had a statistically significant increase in maximum urinary flow rate compared to placebo, at the second and at the fourth month of treatment, and this statistically significant increase in maximum urinary flow rate was maintained from month 4 through month 12. At 1 year in the North American study, a 1.5-mL/s increase for the group treated with finasteride was observed (15). In the open extension study, the men who had been on treatment for 3 years averaged a 2.3-mL/s increase in maximum urinary flow rate (16).

With respect to symptoms, it is accepted that BPH is a clinical disorder that causes an impaired quality of life in many men with symptomatic BPH. A small number of men with BPH develop significant, and at times, irreparable damage to their bladders and kidneys, but that is the exception rather than the rule. The symptoms of BPH can be grouped into two categories, obstructive and irritative (5, 6). Obstructive symptoms include a weak stream, intermittency, dribbling, hesitancy, and incomplete voiding. The irritative symptoms include nocturia, urinary frequency, urgency, and dysuria.

Figure 8 shows that the patients treated with finasteride had a greater decrease in total symptom score than the patients treated with placebo. This decrease became statistically significant compared with placebo during months 6 through 12 in the North American study (15). In the international studies, the decrease in total symptom score compared with placebo was statistically significant at an earlier time point. For patients

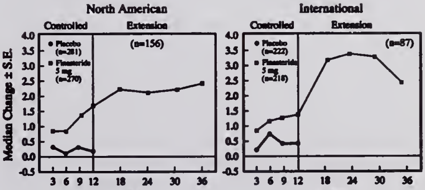


Fig. 7. Effect of finasteride compared with placebo on maximum urinary flow rate over 36 months of treatment: mean change from baseline (±S.E.). (From Stoner et al. (16).)

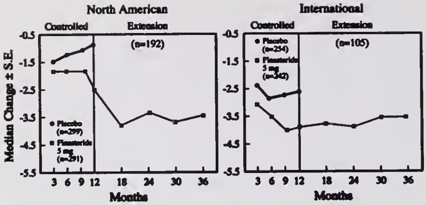


Fig. 8. Effect of finasteride on BPH symptoms over 36 months. Total symptom score: mean change from baseline (±S.E.). (From Stoner et al. (16).)

who received placebo, the data reached a plateau after initial improvement. When patients were followed up for 3 years, the total symptom score decreased by 3.6 points (16). In addition, both patients and investigators were asked in blinded fashion whether they thought that their urologic symptoms were better at the end of 1 year. Both the patients treated with finasteride and the investigators answered more favorably about the treated patients than the placebo group. In a recently completed 2-year study of patients with moderately symptomatic BPH, improvement of symptoms in finasteride-treated patients was maintained in the second year whereas placebo patients returned to baseline severity or deteriorated further in the second year (18).

A urodynamics study carried out in Finland provided important data on finasteride (19). This study evaluated 36 men, 18 of whom were on drugs for 6 months and 18 on placebo. The detrusor opening pressure, detrusor pressure at maximum flow, and the maximum detrusor pressure in the drug-treated group all improved significantly compared with placebo. The data emphasize that shrinkage of the prostate can be associated with urodynamic changes suggesting relief of physical obstruction. Larger scale urodynamic studies are in progress to confirm the smaller study.

A new drug, no matter how well tolerated it appears in clinical studies, must be evaluated over time. It is only after large numbers of people have received the drug in clinical trials, as well as in postmarketing surveillance, that one can begin to assess long-term safety and tolerability.

The evidence so far shows that finasteride has been well tolerated, and the drop-out rate for adverse events in clinical studies has been low. The only adverse experiences observed that were statistically significant compared with placebo in the 12-month controlled clinical studies were related to sexually related events—impotence, decreased libido, and decreased volume of ejaculate in a small percentage of patients (15). The adverse experience profile for patients who received finasteride during the 2 years after the 12-month placebo-controlled clinical study showed no evidence of an increase in the incidence of adverse experiences with increased duration of treatment with finasteride (16). About two thirds of the men who reported sexual adverse experiences had a reversal of these events despite continued treatment with finasteride. In marketing experience, breast tenderness/enlargement and hypersensitivity reactions such as lip swelling and skin rash have been reported.

Because finasteride decreases the level of PSA, information pertaining to the diagnosis of prostate cancer in patients treated with finasteride will be reviewed. Prostate cancer is an important disorder that has received a great deal of attention. PSA, a serine protease, is found in prostate epithelial cells, whether benign or malignant, but PSA, although prostate specific, is not specific for prostate cancer. Although the accepted upper limit of normal for PSA is 4 ng/mL in most assays, some men have prostate cancer despite a PSA in the normal reference range. A number of men who have PSA levels between 4 ng/mL and 10 ng/mL have prostate cancer, much of it organ-confined; and prostate cancer is very likely when the level exceeds 10 ng/mL. Recent studies have proposed that age-specific reference ranges for PSA, PSA density, and PSA velocity may have the potential for increasing sensitivity of detection in younger men whereas increasing specificity in older men (20).

During the 1-year placebo-controlled large-scale studies of finasteride, the number of patients who were detected with prostate cancer was small, but the numbers were comparable in each of the three treatment arms, that is, placebo, finasteride 1 mg, and the finasteride 5 mg groups (15). Finasteride decreases the level of PSA in the serum, but it does so in a highly predictable fashion (20). PSA values fall approximately 50% and stabilize at a new baseline within 6 to 12 months. **Figure 9** shows the cumulative distribution of PSA values at baseline and at month 12 among men with BPH (but without known prostate cancer) who received finasteride 5 mg for 12 months in the North American study. Comparison of the curves at baseline and at month 12 shows that for the same percentile, the serum PSA at month 12 is approximately one half the corresponding value on the baseline curve.

Practically speaking, it is important that a baseline PSA and digital rectal examination be performed in patients before beginning therapy with finasteride, and periodically thereafter. In most patients a rapid decrease in the PSA level is seen within the first few months of therapy, after which the PSA levels stabilize at the lower baseline.

In a typical patient being treated with finasteride for 6 months or more, the PSA value should be doubled when comparing it with normal PSA reference ranges. It is important to note that although the PSA may be within the normal range, a patient may still have prostate cancer, as is true for the general population of men (i.e., who are not on finasteride) who, even with a normal PSA, may have prostate cancer. Guess et al. (21) concluded that the ability to distinguish between BPH and prostate cancer using PSA values does not appear to be adversely affected by treatment with finasteride.

Perspective

BPH is a complex disorder that includes symptoms, reduced urine flow rate, prostatic enlargement, and bladder dysfunction. We know much but not nearly enough about the interactions of various factors in this disorder (22). In addition to looking at the results of medical and surgical therapy, we need to learn more about the natural history of the disease from placebo-controlled studies of intervention.

In the past, only two therapeutic choices were available to physicians and patients—watchful waiting or surgery. Wennberg's classic studies (23) in 11 different communities showed that prostatectomy rates varied almost fourfold and depended on symptoms, patients' desire for treatment, and the opinion of the urologist. A variety of new urologic procedures such as laser therapy, thermotherapy, urethral stents, transurethral needle ablation, high-intensity focused ultrasound, and other approaches are now being studied, but long-term data are not yet available. Transurethral resection of the prostate is still the gold standard for surgical treatment of this disease. There are clear indications for surgery in men with recurrent and/or advanced obstruction, hematuria, or bladder stones. However, we are now in a new era with medical therapy available. Published BPH guidelines suggest that men with moderate to severe symptoms be offered all the choices (medical or surgical therapy or watchful waiting), and be significantly involved in the decision-making process (24).

What about medical therapy? There are two kinds. One is 5-alpha-reductase inhibition (as described in this review) and the second alpha blockers. Why alpha blockade? There are smooth muscle cells in the stromal tissue of the prostate and the bladder neck that can be increased in tone, and it has been shown that alpha blockers relieve this smooth muscle contraction (24). Alpha blockers thus can provide improvement in urinary flow and symptoms in some men with symptomatic BPH. However, alpha blockers (originally developed as antihypertensive drugs) may have significant vascular side effects when used in the treatment of BPH.

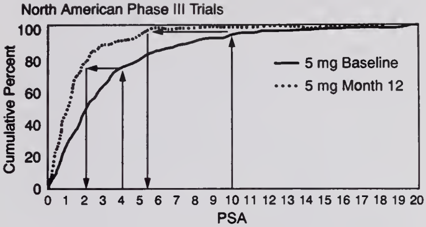


Fig. 9. Effect of finasteride on serum (PSA) at baseline and month 12. PSA cumulative percentile plot. (From Guess et al. (21).)

Fortunately, there are now more choices for men and a new opportunity exists. Internists and other primary care physicians can broaden their diagnostic and therapeutic horizons and also work closely with urologists. A primary care physician who completes a careful history and physical examination and uses the standardized and validated questionnaire can facilitate the diagnosis of BPH and intervene medically in appropriate patients (25). There are clearly times when a urologist's opinion is needed, particularly when there is suspicion of cancer or the PSA is elevated, but in many instances, the management can and should be carried out by the primary care physician. Urologists can also help internists and other primary care physicians to perform better rectal examinations. There is an excellent opportunity for both groups to work closely together, and the ones who will benefit from this kind of cooperative effort are older males.

Questions and Answers

Question: Do you have any data on morbidity and mortality as affected by the drug? What is the cost of the drug?

Sherwood: BPH does not generally lead to mortality, but in the published literature on prostatectomy, there is a small, but definite incidence of mortality around 1%. Depending on the pharmacy, the cost of finasteride is about $600 a year.

Question: Any data on acid phosphatase levels?

Sherwood: Acid phosphatase is much less specific than the PSA, and most physicians do not focus attention on acid phosphatase at this point. Some urologists find it useful in following up patients with known prostate cancer in terms of progression. Measurement of PSA is generally viewed as much more useful.

Dr. Marvin Levitt: How do you quantify symptoms, particularly nocturia in elderly people who may be affected with heart failure, hypertension, or kidney disease?

Sherwood: As far as symptoms go, there has been an attempt to rationalize the evaluation of BPH. The American Urological Association has validated a questionnaire that you can use in your office for patients (25) (see **Fig. 10**). It consists of seven questions, and it correlates well with BPH.

Dr. Michael Droller: I would like to speak from a medical point of view. As Dr. Sherwood commented, during his overview of this problem, it is a complex issue. BPH is one term that has been used interchangeably for histologic changes, for physical examination changes, for symptom changes, and represents many aspects and different disease processes that result in a patient seeking some assistance. There have been no correlations in urodynamic evaluations with either symptom findings or outcome from a variety of treatments. That is why the studies that Dr. Sherwood is describing will be important. Symptom scores have also been the subject of a great deal of discussion and controversy. We are just at the beginning of starting to sort out what this disease complex represents and the variety of different agents to treat a variety of different approaches. The systematic study of those agents is going to be critical over the next several years as we learn more about it. I would urge those with questions about patients with symptoms that they evaluate them carefully for the specificity of the disease process that has led to those symptoms.

Dr. Paul Gilbert: Thank you. My fear was that when Dr. Sherwood started talking about risk factors, he spoke of risk factors of smoking for heart disease and cholesterol for heart disease. I was afraid that somebody was going to say sexual activity was a risk factor for prostatic disease.

Sherwood: Sexual activity is not known to be a risk factor for BPH.

BPH:
Obstructive and Irritative Symptoms

	Not at all	Less than 1 time in 5	Less than half the time	About half the time	More than half the time	Almost always
• 1. Incomplete emptying Over the past month, how often have you had a sensation of not emptying your bladder completely after you finished urinating?	0	1	2	3	4	5
• 2. Frequency Over the past month, how often have you had to urinate again less than two hours after your finished urinating?	0	1	2	3	4	5
• 3. Intermittency Over the past month, how often have you stopped and started again several times when you urinated?	0	1	2	3	4	5
• 4. Urgency Over the past month, how often have you found it difficult to postpone urination?	0	1	2	3	4	5
• 5. Weak Stream Over the past month, how often have you had a weak urinary stream?	0	1	2	3	4	5
• 6. Straining Over the past month, how often have you had to push or strain to begin urination?	0	1	2	3	4	5
• 7. Nocturia Over the past month, how many times did you most typically get up to urinate from the time you went to bed at night and until the time you got up in the morning?	0	1	2	3	4	5

Fig. 10. AUA validated symptom index for BPH. (From Barry et al. (25).)

Acknowledgment

The author wishes to express his appreciation to Dr. Leonard Silverstein for his input and helpful review of the manuscript.

References

1. Guess HA. Benign prostatic hyperplasia antecedents and natural history. Epidemiol Rev 1992; 14:131–135.
2. Berry SJ, Coffey DS, et al. The development of human benign prostatic hyperplasia with age. J Urol 1984; 132:474–479.
3. Coffey DS, Walsh PC. Clinical and experimental studies of benign prostatic hyperplasia, Urol Clin North Am 1990; 17:461–475.
4. Carter HB, Pearson JD, et al. Longitudinal evaluation of prostate-specific levels in men with and without prostate disease. JAMA 1992; 267:2215–2220.
5. Chute CG, Panser LA et al. The prevalence of prostatism: a population-based survey of urinary symptoms. J Urol 1993; 150:85–89.
6. Girman CJ, Panser LA et al. Natural history of benign prostatic hyperplasia: urinary flow rates in a population-based study. J Urol 1993; 150:887–892.
7. Imperato-McGinley J, Guerrero L et al. Steroid 5 alpha-reductase deficiency in man: an inherited form of male pseudohermaphroditism. Science 1974; 186:1213–1215.
8. Wilson JD, Harrod MJ et al. Familial incomplete male pseudohermaphroditism, type 1. N Engl J Med 1974; 290:1097–1103.
9. Gabrilove JL, Levine AC et al. Effect of long-acting gonadotropin-releasing hormone analog (leuprolide) therapy on prostatic size and symptoms in 15 men with benign prostatic hypertrophy. J Clin Endocrinol Metab 1989; 69:629–632.
10. Peters CA, Walsh PC. The effect of nafarelin acetate, a luteinizing-hormone-releasing hormone agonist, on benign prostatic hyperplasia. N Engl J Med 1987; 317: 599–604.
11. McCornell JD. Benign prostatic hyperplasia: hormonal treatment. Urol Clin N Am 1995; 22:387–400.
12. Peterson RE, Imperator-McGinley J et al. Male pseudohermaphroditism due to steroid 5 alpha-reductase deficiency. Am J Med 1977; 62:170–191.
13. Thigpen AE, Davis DL et al. Molecular genetics of steroid 5 alpha-reductase 2 deficiency. J Clin Invest 1992; 90: 799–809.
14. Stoner E. The clinical development of a 5-alpha reductase inhibitor, finasteride. J Steroid Biochem Mol Biol 1990; 37:375–378.
15. Gormley GJ et al. The effect of finasteride in men with benign prostatic hyperplasia. N Engl J Med 1992; 327: 1185–1191.
16. Stoner E and the Finasteride Study Group. Three-year safety and efficacy data on the use of finasteride in the treatment of benign prostatic hyperplasia. Urology 1994; 43:284–294.
17. Tempany CMC, Parten AW et al. The influence of finasteride on the volume of the peripheral and periurethral zones of the prostate in men with benign prostatic hyperplasia. Prostate 1993; 22:39–42.
18. Andersen JT, Ekman P et al and the Scandinavian BPH study group. Can finasteride reverse the progress of BPH? A two-year placebo-controlled study. Urology In press.
19. Tammella TLJ, Kontturi MJ. Urodynamic effects of finasteride in the treatment of bladder outlet obstruction due to benign prostatic hyperplasia. J Urol 1993; 149: 342–344.
20. Oesterling JE, Jacobsen SJ et al. Serum prostate-specific antigen in a community-based population of healthy men. JAMA 1993; 270:860–864.
21. Guess HA, Heyse JF et al. The effect of finasteride on prostate-specific antigen in men with benign prostatic hyperplasia. Prostate 1993; 22:31–37.
22. Girman CJ, Panser LA et al. Natural history of prostatism: relationship among symptoms, prostate volume and peak urinary flow rate. J Urol 1995; 153:1510–1515.
23. Wennberg JE, Mulley AG et al. An assessment of prostatectomy for benign urinary tract obstruction. JAMA 1988; 259:3027–3030.
24. McConnell J, Barry M, Brushewitz R et al. Benign prostatic hyperplasia diagnosis and treatment, Quick reference guide for clinicians, No. 8, AHCPR publication no. 94-0583. Rockville, MD: Agency of Health Policy and Research, Public Health Services, U.S. Department of Health and Human Services, 1994.
25. Barry MJ, Fowler FJ et al. and the Measurement Committee of the American Urological Assoc. The AUA symptom index for BPH. J Urol 1992; 148:1549–1557.

Thyroid Nodules and the Detection of Thyroid Cancer

Richard S. Haber, M.D.

Abstract

Although thyroid nodules are common and cause concern about possible malignancy, relatively few are thyroid cancers, and relatively few thyroid cancers are lethal. These considerations justify a selective approach in recommending surgical excision of nodules. Selection of patients should be based on assessment of cancer risk factors in the history and physical examination, and most importantly on fine-needle aspiration biopsy of the thyroid with cytologic examination. Other testing in most cases should be limited to serum thyroid-stimulating hormone and antithyroid antibodies and a thyroid sonogram. With the results of this evaluation a recommendation can be made for either immediate surgery or observation. This approach should reduce the number of operations for benign thyroid nodules and avoid excessive diagnostic testing.

The discovery of a thyroid nodule by a physician or patient usually leads to apprehension about possible thyroid cancer. Although the majority of these nodules are harmless, even after extensive medical testing there is typically residual uncertainty about a nodule's malignant potential, frequently leading to thyroid surgery for benign disease. In which patients are the chances that a nodule is malignant high enough to justify surgical excision? We will review herein the problem of thyroid nodules, and suggest a current approach to their evaluation and management.

Epidemiology of Thyroid Nodules and Cancer

The prevalence of palpable thyroid nodules in the adult U.S. population is in the range of 4% to 7% (1, 2), and there is a strong female predominance (4:1 female/male ratio) (3). The annual incidence is about 0.1% (3), or about 250,000 persons per year. This compares with 12,000 new cases of thyroid cancer per year, and about 1000 thyroid cancer deaths (2). These figures suggest that only about 5% of palpable thyroid nodules are cancers. The somewhat higher 8% to 17% cancer rate in surgically excised nodules (2) presumably reflects the selection of suspicious cases for surgery. In addition, overall mortality from thyroid cancer is low, about 9%. Certainly these statistics argue against the indiscriminate surgical excision of most thyroid nodules.

As common as palpable nodules may be, nonpalpable thyroid nodules are even more common. Autopsy and ultrasound surveys of the general population show that the prevalence of these small nodules (usually <1 cm) increases with age, exceeding 40% after age 50 (2). Although small "subclinical" thyroid cancers are also common in autopsy studies, it is clear that exceedingly few small nonpalpable nodules pose any threat to the patient's well-being (1).

Pathology of Benign Thyroid Nodules

Most thyroid nodules are either colloid nodules or follicular adenomas. *Colloid nodules* consist histologically of enlarged and variably sized follicles, as are found in multinodular goiter. They may be recognizable clinically when they comprise a multinodular goiter, but can also

---

From the Department of Medicine, Division of Endocrinology, Mount Sinai School of Medicine, New York, New York. Address reprint requests to Richard S. Haber, M.D., Mount Sinai School of Medicine, Box 1055, One Gustave L. Levy Place, New York, NY 10029-6574.

TABLE 1
Types of Thyroid Cancer

Type of cancer	Comments
Papillary carcinoma	70% of thyroid cancers; usually good prognosis; characteristic nuclear changes facilitate cytologic diagnosis; usual form in younger patients
Follicular carcinoma	10–15% of thyroid cancers; prone to distant metastases with associated mortality; older age group
Anaplastic carcinoma	5% of thyroid cancers; highly lethal; older age group
Medullary carcinoma	5–10% of thyroid cancers; sporadic or inherited; may be detected by elevated serum calcitonin
Others	Lymphoma (<5%), metastatic carcinoma (rare)

present as apparently solitary nodules, which may be classified as "macrofollicular adenoma." The cause of these nodules is obscure in most cases, although nodular goiter frequently is familial. In other parts of the world (but not in the United States), endemic iodine deficiency is a major cause of nodular goiter.

Follicular adenomas are typically solitary lesions, consisting histologically of uniformly sized follicles with a well-defined capsule. They are thought to be clonal neoplasms arising from mutation(s) in a single cell. Most are hypofunctioning in terms of iodide uptake and thyroid hormone synthesis, but a few show increased uptake on radioiodine scanning ("hot nodules") and function autonomously of thyroid-stimulating hormone (TSH).

With sufficient growth, some of these solitary autonomous nodules may cause thyrotoxicosis. Somatic point mutations in the gene for the TSH receptor are known to cause many of these solitary autonomous nodules (4). The mutations cause the receptor to constitutively activate thyroid cell adenyl cyclase independently of TSH, stimulating both growth and function of the affected cells.

Hashimoto's (chronic lymphocytic) thyroiditis, although typically causing diffuse goiter, may cause thyroid nodules, as can subacute (painful) thyroiditis. *True cysts* are rare; most thyroid "cysts" represent degeneration within solid nodules, either benign or malignant. An exception is the *thyroglossal duct cyst,* a midline remnant of embryonic thyroid migration. *Thyroid hemiagenesis* is rare, but causes compensatory enlargement (pseudonodule) of the contralateral (usually right) lobe (5).

Pathology of Thyroid Cancer

Types of thyroid cancer are shown in **Table 1.** *Papillary carcinoma,* the most common form, is fortunately the most indolent. Because it does not usually cause distant metastases, most cases are cured by thyroid surgery. This tumor has a propensity to metastasize to cervical lymph nodes, but the presence of local lymph node metastases per se does not alter the good prognosis, and local neck disease can usually be controlled by a combination of surgery and/or radioiodine therapy. A subset of more aggressive papillary carcinomas may cause significant morbidity and mortality by invasion of adjacent neck tissues, usually in older patients. In addition to age (>40 years), negative prognostic factors include penetration of the thyroid capsule by the primary tumor, and size of the primary tumor (6). Occasionally there are distant metastases, especially to the lung.

Follicular carcinomas are less common, but are prone to metastasize to distant sites (lung and bone), and are then associated with significant mortality (50% or more at 5 years) (6). The Hurthle cell variant may be especially aggressive. *Anaplastic thyroid carcinoma* is mainly seen in older patients and presents as a rapidly growing, sometimes painful mass. Survival is measured in months. *Medullary carcinoma of the thyroid,* arising from the C cells, is most often sporadic (80% of cases) but can occur as an inherited autosomal dominant disease, either without involvement of other endocrine glands (familial medullary thyroid carcinoma), or together with pheochromocytoma and/or hyperparathyroidism (multiple endocrine neoplasia type II). Most familial cases are caused by inherited point mutations in the *ret* proto-oncogene (7), which can be detected by DNA sequence analysis on blood samples. Ten-year survival is about 60%. *Lymphoma of the thyroid* typically occurs against a background of Hashimoto's thyroiditis and is more common with increasing age.

Radiation and Thyroid Cancer

From the 1930s to 1970s about a million American children received external x-ray treatments for a variety of benign diseases, including acne, tinea capitis, and enlarged tonsils or thymus glands. These treatments delivered a low dose of radiation to the thyroid (10–200 rads). It has been shown repeatedly that in children who received neck irradiation for benign disease there is a markedly increased incidence of thyroid nodules, and that an unusually high proportion of these nodules are thyroid cancer (8). In one large series the incidence of clinically evident thyroid nodules in these irradiated subjects reached 39% after long-term follow-up for decades. Moreover, one third of the nodules that were surgically excised proved to be malignant (9).

The explosion of the nuclear reactor at Chernobyl in 1986 released large amounts of radioiodine, irradiating the thyroids of children in Ukraine and nearby Belarus (Byelorussia). It has recently become evident that the disaster has caused a large increase in the incidence of thyroid cancer in these children. As with therapeutic neck irradiation, reports of increased thyroid cancer have been restricted thus far to individuals who were irradiated in childhood. It thus appears that the thyroid in childhood is particularly prone to the carcinogenic effects of ionizing radiation.

Clinical Approach to Thyroid Nodules

The relatively low incidence of clinically significant thyroid cancer as well as the indolent behavior of the majority of thyroid cancers argue against indiscriminate recommendations for thyroid surgery to patients with nodules. Instead, the physician should make a judgment on the relative likelihood that a nodule is malignant based on clinical findings, recommending surgery for more suspicious nodules and observation for others. The use of fine-needle aspiration biopsy and the advent of high-resolution thyroid sonography have altered the traditional approach based on radionuclide scanning (see below).

History

As noted above, it is critical to inquire about head and neck irradiation in childhood. Age and gender are also important factors. The high incidence of benign nodules in women in the middle years of life lowers the likelihood that a nodule is malignant in this group. Conversely, suspicion of cancer should be higher in men at any age, and in both sexes after age 60 or before age 20. Moreover, thyroid cancers in older patients have greater malignant potential (6).

A family history of multinodular goiter or of Hashimoto's disease and hypothyroidism is suggestive of a benign etiology. Thyroid cancer is not usually familial. Exceptions are familial medullary carcinoma, with or without multiple endocrine neoplasia type II, and familial colonic polyposis, which is associated with thyroid cancer.

Symptoms such as rapid growth, dysphagia, and hoarseness are suggestive of carcinoma, but are absent in most cases of thyroid cancer. A painful nodule is usually due to subacute thyroiditis or to hemorrhage into a benign nodule, but can occur in the uncommon anaplastic carcinoma.

Physical Examination

The principal distinction to be made on physical examination is between a generalized nodular goiter (multinodular goiter) and an apparently solitary nodule. Because the incidence of cancer in multinodular goiter is low in the United States, observation alone is sufficient for such patients, with further evaluation reserved for cases of suspicious growth in a "dominant" nodule. It is the apparently solitary nodule that should raise the most concern. The finding of one or more additional small nodules by sonography should not alter the initial characterization of a nodule as "solitary," because nonpalpable nodules are common in the general population and are associated with both benign and malignant nodules.

Although thyroid cancers are typically very firm or hard in consistency, neither consistency nor size has proved to be a reliable indicator of malignancy. However, fixation of the nodule to surrounding tissues or associated cervical adenopathy are highly suspicious findings.

Laboratory Evaluation

Blood tests have a limited role. Serum TSH should be measured routinely by a sensitive assay. A suppressed TSH suggests the presence of a hyperfunctioning nodule, which can be confirmed with radioiodine scanning (see below). An elevated TSH indicates hypothyroidism and suggests the presence of Hashimoto's thyroiditis.

Measurement of serum antithyroid antibodies (antithyroid peroxidase and antithyroglobulin) may be useful because high titers suggest a diagnosis of Hashimoto's thyroiditis. Measurement of serum calcitonin to detect medullary carcinoma of the thyroid should be performed when this disease is suspected based on family history, or the presence of chronic diarrhea. Its diagnostic value in other nodule patients is unclear at present. However, serum calcitonin is probably more likely to be elevated in men with nodules than in women, because the overall number of nodules but not the incidence of medullary carcinoma is lower in men. Serum thyroglobulin, a useful marker for detecting recurrence of thyroid cancer in thyroidectomized patients, is elevated

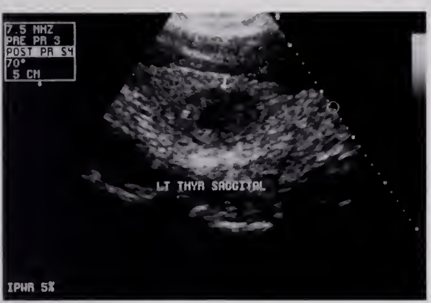


Fig. 1. Thyroid sonogram in a 37-year-old man showing sagittal view of a hypoechoic nodule measuring 1.5 cm in length. The nodule (arrow), which was detected incidentally during carotid artery sonography for unrelated symptoms, was not easily palpable due to obesity. Fine-needle aspiration biopsy under sonographic guidance yielded a diagnosis of medullary carcinoma of the thyroid.

in both benign and malignant thyroid disease and is not useful in the evaluation of nodules.

Imaging Studies

High-resolution *thyroid sonography* has replaced radionuclide thyroid scanning as the procedure of choice for imaging thyroid nodules. Thyroid sonography provides a precise and reproducible measurement of nodule size, which is essential for accurately following subsequent growth or shrinkage of a nodule. In many patients in whom accurate palpation is hindered by obesity or kyphosis, sonography provides the only accurate delineation of a nodule (**Fig. 1**). In addition, sonography demonstrates whether a nodule is cystic, solid, or mixed ("complex"). In the occasional nodule that is purely cystic, carcinoma is unlikely. Although the finding that a nodule is complex is not helpful in diagnosis, identifying the cystic component allows recognition of fluid accumulation that could otherwise be mistaken for growth of solid tissue.

Identification of sonographic characteristics that would be useful in the diagnosis of thyroid cancer has proved elusive. However, sonography can identify the presence of a diffusely abnormal gland (due to nodular goiter or Hashimoto's thyroiditis) that might not be detected on physical examination in some patients, and thus point to a benign etiology.

Radionuclide (radioiodine/pertechnetate) scanning of the thyroid has limited usefulness despite its traditional central role in the evaluation of nodules. Scanning provides a measure of the iodine-trapping function in a nodule compared to the surrounding thyroid tissue. Nodules may thus be classified as "cold," "warm," or "hot." Most are "cold" (decreased uptake, 85%) or "warm" (uptake similar to surrounding tissue, 10%) including both cancers and benign nodules (10, 11). Only the finding of a "hot" nodule (increased uptake with suppression of uptake in extranodular tissue), occurring in only 5% of cases, provides helpful information. "Hot" nodules usually represent autonomous hyperfunctioning adenomas and are almost always benign. Because the occasional hyperfunctioning nodule can now be detected with modern sensitive serum TSH assays (suppressed TSH) and/or by thyroid sonography (atrophic contralateral lobe), routine use of thyroid scanning in the evaluation of nodules is wasteful and inefficient. I reserve this test for those cases in which hyperfunctioning nodules are suspected based on the presence of thyrotoxicosis and/or suppressed TSH. It need not be performed routinely.

Fine-Needle Aspiration Biopsy of the Thyroid

Fine needle aspiration of the thyroid is currently the most useful test in the evaluation of nodules. Although it is still not possible to exclude thyroid cancer with certainty without thyroid surgery, the use of fine-needle aspiration biopsy and cytology over the past two decades has made it possible to avoid surgery in many if not most patients with thyroid nodules.

The procedure is easily performed in the office. With the patient supine and the neck extended, the skin is penetrated with a 22-gauge needle attached to a 10-mL syringe. When the needle enters the nodule, negative pressure is applied while the needle is moved up and down to disrupt the tissue. When bloody material appears in the needle hub, negative pressure is released and the needle is withdrawn. The aspirated mixture of blood and thyroid (ideally one drop) is smeared between two glass slides, which are immediately sprayed with ethanol fixative, and sent for staining and cytologic interpretation. Multiple needle passes in different parts of the nodule should be done to avoid sampling error.

In many cases the cytologic review will yield a benign or malignant diagnosis. In these situations the choice between observation and surgery is not difficult. The false negative rate for benign diagnoses in recent reported series is 1% to 10%, and most thyroid cancers are detected (80% to 90%) (12). Fine-needle aspiration biopsy thus has the potential to correctly select patients with cancer for surgery and to reassure many others that they should be observed.

However, the test has several weaknesses. First, the accuracy of the test requires skill and experience in both the person performing the biopsy as well as in the cytologist interpreting the smears. Ideally the procedure should be done by someone who performs it frequently (at least 25 cases per year). Second, inadequate specimens consisting mainly of blood with very few thyroid follicles are common, even in experienced hands (about 20%). Third, even with adequate specimens, the cytologists may report a microfollicular aspirate, and then be unable to distinguish between microfollicular adenoma and well-differentiated follicular carcinoma. This indeterminate result also occurs in about 20% of cases. Fourth, the occurrence of at least some false negative results means that thyroid cancer can never be excluded with absolute certainty.

A refinement to the technique of fine-needle aspiration biopsy is provided by real-time sonographic guidance, when available. This method has several advantages. One can confirm that the needle is indeed within the nodule and not in adjacent tissue. This should reduce false negative results. In addition, nonpalpable nodules can be easily aspirated when appropriate (Fig. 1). Moreover, the chances of obtaining adequate specimens can be improved by avoiding areas of cystic degeneration within nodules and by directing the needle to solid portions of complex nodules (Fig. 2).

A summary of recommended diagnostic testing for solitary thyroid nodules is given in Table 2. In most cases, testing should be limited to serum TSH, thyroid sonography, and fine-needle aspiration biopsy.

Thyroid Hormone Suppression Therapy

Another traditional method for evaluating nodules is suppression therapy with L-thyroxine. This approach is based on the concept that benign nodules but not cancers are dependent on TSH and that suppressing pituitary TSH secretion with exogenous thyroid hormone will selectively shrink benign nodules. The diagnostic usefulness of this maneuver is limited by the fact that most nodules, whether benign or malignant, do not shrink with suppression therapy (13). Nonetheless, a subset of benign nodules may shrink during L-thyroxine treatment, providing evidence for a benign diagnosis. L-Thyroxine treatment may also be used in conjunction with observation for growth of a nodule, because growth during suppression therapy is considered suspicious for malignancy. However, there has been concern that long-term suppression of serum TSH below the lower limit of normal, particularly in postmenopausal women, may exacerbate bone loss.

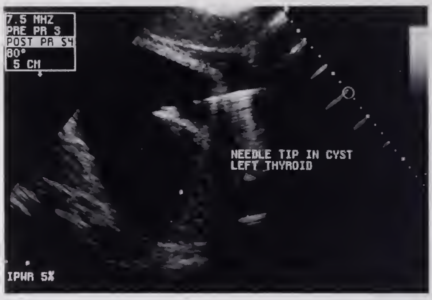


Fig. 2. Thyroid sonogram in a 43-year-old man showing transverse view of a complex nodule with a biopsy needle in view. Sonographic guidance facilitates biopsy of solid components of complex nodules, as well as drainage of cyst fluid (anechoic area).

Who Should Have Thyroid Surgery?

The decision is not always straightforward. Clearly patients with suspicious or malignant cytologic diagnoses should be referred to an experienced thyroid surgeon, and patients with clearly benign cytologic diagnoses in general should be observed. But what of the numerous patients with inadequate or indeterminate results, even after repeat biopsy? In such cases a judgment needs to be made that considers the relative chances of malignancy as well as the patient's ability to tolerate uncertainty. Risk factors such as male sex

TABLE 2
Recommended Diagnostic Testing for Solitary Thyroid Nodules

Test	Comments
Serum TSH	If suppressed, confirm suspected hyperfunctioning nodule with radionuclide scanning
Antithyroid antibodies	May provide evidence for Hashimoto's thyroiditis
Thyroid sonography	Defines nodule size and cystic components; detects unsuspected diffuse thyroid disease
Fine-needle aspiration	Repeat if inadequate specimen; may yield indeterminate result if microfollicular pattern present

and age greater than 60 may argue for surgery, whereas in women aged 30 to 50 years the high prevalence of benign nodules and the relative indolence of most thyroid cancers in this age group might argue for observation with periodic assessment of nodule growth. In the latter case, surgery is recommended for those nodules that grow. Often the patient makes the decision independent of the physician's recommendation, based on fear of surgery on the one hand or fear of cancer on the other.

We recently have reviewed the final histologic diagnoses in 87 consecutive specimens from thyroid surgery performed for nodular disease at Mount Sinai Hospital in March to September 1994. Even with the widespread use of fine-needle aspiration biopsy and cytology to select patients for surgery, only 17% of cases were malignant. In retrospect, most of these operations had no health benefit. It is thus clear that improved diagnostic techniques, such as molecular markers for cancer in fine-needle aspirates, are needed to separate benign from malignant thyroid nodules.

References

1. Rojeski MT, Gharib H. Nodular thyroid disease: evaluation and management. N Engl J Med 1985; 313:428–436.
2. Mazzaferri EL. Management of a solitary thyroid nodule. N Engl J Med 1993; 328:553–559.
3. Vander JB, Gaston EA, Dawber TR. The significance of nontoxic thyroid nodules: final report of a 15 year study of the incidence of thyroid malignancy. Ann Intern Med 1968; 69:537–540.
4. Van Sande J, Parma J, Tonacchera M, et al. Somatic and germline mutations of the TSH receptor gene in thyroid diseases. J Clin Endocrinol Metab 1995; 80:2577–2585.
5. Hamburger JI, Hamburger SW. Thyroidal hemiagenesis. Arch Surg 1970; 100:319–320.
6. Mazzaferri E. Carcinoma of follicular epithelium: radioiodine and other treatments and outcomes. In: Braverman LE, Utiger RD, eds. The Thyroid. 6th ed. Philadelphia: J.B. Lippincott, 1991: 1138–1165.
7. Wells SA. New approaches to the patient with medullary carcinoma of the thyroid gland. Thyroid Today 1994; 17:1–9.
8. Schneider AB. Carcinoma of follicular epithelium: pathogenesis. In: Braverman LE, Utiger RD, eds. The Thyroid. 6th ed. Philadelphia: J. B. Lippincott, 1991: 1121–1129.
9. Schneider AB. Thyroid nodules following childhood irradiation: a 1989 update. Thyroid Today 1989; 12:1–7.
10. Ashcraft MW, Van Herle AJ. Management of thyroid nodules. I. history and physical examination, blood tests, x-ray tests, and ultrasonography. Head Neck Surg 1981; 3:216–230.
11. Ashcraft MW, Van Herle AJ. Management of thyroid nodules. II. scanning techniques, thyroid suppressive therapy, and fine needle aspiration. Head Neck Surg 1981; 3:297–322.
12. Hamburger JI, Kaplan MM, Husain M. Diagnosis of thyroid nodules by needle biopsy. In: Braverman LE, Utiger RD, eds. The Thyroid. 6th ed. Philadelphia: J. B. Lippincott, 1991: 544–559.
13. Cooper DS. Thyroxine suppression therapy for benign nodular disease. J Clin Endocrinol Metab 1995; 80:331–334.

Grand Rounds

The Amyloidoses

Joel Buxbaum, M.D.

Amyloid deposition and its associated diseases, the amyloidoses, are the subject of today's Medical Grand rounds. Although the immunoglobulin (Ig)-related amyloids are the most common form of systemic amyloid deposition seen in clinical practice in the United States, Ig light (L) and heavy (H) chains represent only two of the 15, thus far identified, precursor proteins that produce tissue deposits sufficient to compromise organ function and display the defining qualities of amyloid (i.e., Congo Red binding with positive birefringence under polarized light and the characteristic fibrils seen with the electron microscope). Hence there are now 15 known distinct amyloidoses, with potentially more to come.

History

In the century that elapsed after the original pathologic description of amyloidosis, similar deposits were reported in a score of additional clinical entities. The amyloid-associated disorders now include neoplastic diseases (multiple myeloma, insulinomas, medullary carcinoma of the thyroid), chronic infectious diseases (including leprosy, osteomyelitis, and tuberculosis), environmental insults (the intracutaneous self-administration of heroin), presumably degenerative neurologic disorders (Alzheimer's disease), iatrogenic conditions (hemodialysis-associated amyloid deposition), and genetic diseases (familial Mediterranean fever, familial amyloidotic polyneuropathies) (1). For more than a century, it was not clear whether amyloid represented one or many substances.

Although the tissue deposits were amorphous when examined under the light microscope after hematoxylin-and-eosin staining, ultrastructural studies in the late 1950s and early 1960s demonstrated a discrete fibrillar structure. Subsequent x-ray diffraction studies revealed the pattern characteristic of a β-pleated sheet structure. Chemical analyses of the tissue fibrils were unsuccessful without the development of a reliable method of tissue extraction. Conventional extraction procedures using high salt, extremes of pH, detergents, or proteases did not release amyloid from its tissue matrix. It was not until the discovery that suspending amyloid-laden tissue in distilled water, or in solutions of low ionic strength, allowed almost quantitative recovery of the fibril material as a suspension, which could then be purified and chemically characterized, could the question of one or many amyloids be answered (2). For the past 20 years, many laboratories have used these or analogous procedures to establish the chemical nature of the fibrils in the disparate clinical conditions associated with amyloid deposition.

Chemical Classification

One of the practical difficulties in the chemical analysis of the amyloidoses was ensuring, in the face of such diverse clinical entities, that all the tissue extracts truly contained material with the properties of amyloid, and further, that any putative amyloid substance retained those properties. A definition for amyloid substance, which included a set of tinctorial, ultrastructural, and solubility properties, was established by investigative consensus (**Table 1**). During the past 25 years the chemical nature, and the normal precursors, in 15 different amyloidoses have been discovered (**Table 2**).

Adapted from the author's Grand Rounds presentation at The Mount Sinai Medical Center on April 4, 1995. Address reprint requests to Joel Buxbaum, M.D., Research Service, Department of Veterans Affairs Medical Center, 423 East 23rd Street, New York, NY 10010.

TABLE 1
Histologic Criteria for Definition of Amyloid

1. Homogeneous, hyaline, eosinophilic material on H&E
2. Crystal Violet metachromasia
3. Stains with alkaline Congo Red
4. Apple green positive birefringence with polarized light after Congo Red staining
5. Fibrillar structure on electron microscopy
6. Stains with antibody specific for P-component
7. Stains with antibody directed toward specific precursor

Chemical Classification and Pathologic Diagnosis

The diagnosis of any of the amyloidoses is made by the pathologist. The critical assay is the staining of the tissue sample with Congo Red followed by microscopic examination under polarized light. The amyloid deposits bind the stain, giving a homogeneous pink color with standard illumination, and bright green birefringence when the beam is polarized. False positives and negatives are rare, the latter usually occurring when only small amounts of material are present. With the identification of specific protein precursors, it has become possible to not only identify the presence of amyloid per se, but to determine the precursor, and thus the nature of the disease. In addition, antibody directed against the salt-extractable P-component, found in all forms of amyloid, provides a more definitive assay than Congo Red binding.

Clinically, we do not necessarily sample an involved organ for tissue for pathologic examination, although the risks of such sampling (of kidney, heart, and liver) have declined markedly over the years. It has become clear that, in most forms of systemic amyloid deposition, Congo Red or antibody positive material can be obtained by aspiration of subcutaneous fat or biopsy of the rectal mucosa, including vessels, in 80% to 90% of patients (3). In the remaining 10% the affected organ must be examined. In some instances, even in the face of a positive rectal or subcutaneous fat biopsy, it is necessary to biopsy a potentially involved organ to ensure that the observed dysfunction is actually related to amyloid deposition and not some other coincidental clinical entity.

Table 2 shows the current chemical classification of amyloid. AL, comprising Ig L-chains and their fragments, used to be called primary, but we now recognize that the same deposits are also found in the amyloid associated with multiple myeloma. Thus AL is associated with plasma cell disorders and the clinical distinction is based on the degree of neoplastic growth which characterizes the clone producing the precursor amyloidogenic protein.

AA was always considered to be secondary to some infectious or inflammatory disease and is probably the most common cause of systemic amyloid deposition worldwide. Rheumatoid arthritis is the most frequent stimulus for AA amyloid in this country, occurring in 3% to 5% of patients with the disease, but having clinical significance in only half of those with tissue deposits (4,5). The table probably represents between 90% and 100% of the amyloid precursors that will ultimately be defined.

Pathogenesis

In all forms of amyloid the precursor is synthesized as a soluble protein (**Fig. 1**). Amyloid fibrils are rarely seen inside the synthesizing cell, so it appears that all the soluble precursors are either processed, or encounter conditions in vivo, which result in aggregation, fibril formation, and tissue deposition. One hypothesis proposes that the molecules are intrinsically unstable but are maintained soluble intracellularly by a set of molecules collectively called *chaperonins,* which transports them into the secretory pathway, or, if they are defective, to the lysosomes for degradation into peptides. In many of the amyloidoses the soluble precursors are synthesized at a distance from the site of deposition; thus either the precursor is synthesized in a form that is stable both inside the cell and in the circulation, or is protected from fibrillogenic aggregation by the intracellular chaperonin system and extracellularly by interaction with other molecules. Another hypothesis suggests that fibril precursors may be aberrant folding intermediates of monomers that are generated during the denaturation of the polymeric molecule before lysosomal proteolysis (6). This may be true, but it is not clear that the sites of clinical deposition correspond to the normal sites of precursor catabolism, hence an additional mechanism must be invoked to account for precursor solubility in the circulation.

We do not know if any changes in the precursor take place in the circulation; we know that virtually all these proteins, although synthesized as monomers, when deposited as fibrils are polymers of a basic subunit. It is not clear that the normally assembled circulating polymer is the fibril precursor, or whether a monomer uses another polymerization pathway to form the tissue fibrils. The latter has been suggested to be the case in transthyretin amyloidogenesis (6). It is

TABLE 2
Chemical Classification of the Human Amyloidoses

Amyloid protein	Precursor protein	Clinical syndrome
AL	Immunoglobulin L-chain	Primary amyloid; multiple myeloma
AH	Immunoglobulin H-chain	Primary amyloid; multiple myeloma
AA	apo-SAA	Inflammation-associated familial Mediterranean fever; Muckle-Wells
ATTR	Transthyretin	Familial amyloidotic polyneuropathy; senile systemic amyloidosis; isolated vitreous amyloid
AGEL	Gelsolin	Familial amyloidotic polyneuropathy; Finnish lattice corneal dystrophy
AApoAI	Apolipoprotein AI	Familial amyloidotic polyneuropathy; Van Allen
$A\beta_2M$	Beta-2 microglobulin	Dialysis-associated amyloidosis
A_β	Beta-protein precursor	Alzheimers; Down syndrome; HCHWA (Dutch)
ACys	Cystatin C	HCHWA (Iceland)
ACal	Calcitonin	Medullary carcinoma of the thyroid
AIAPP	Islet amyloid polypeptide	Insulinoma; type II diabetes mellitus
AANF	Atrial natriuretic factor	Isolated atrial amyloid
AScr	Scrapie (prion) protein	Creutzfeldt-Jakob; Gerstmann-Straussler-Scheinker syndrome
AFibrin	Alpha chain fibrinogen	Hereditary renal amyloid
ALys	Lysozyme	Hereditary non-neuropathic systemic amyloidosis (Ostertag type)

HCHWA, hereditary cerebral hemorrhage with amyloidosis.

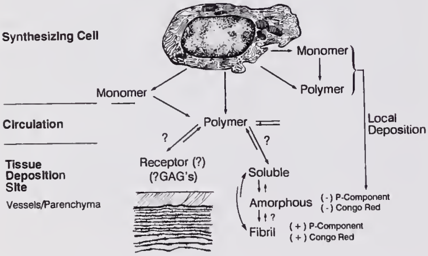


Fig. 1. The amyloid precursor-synthesizing cell produces a soluble monomer product that undergoes normal polymerization within the cell or at the time of secretion. It is unclear whether specific processing or further polymerization occurs within the circulation or at the time of deposition. Although several proteins can be isolated from the deposits by salt extraction, notably the P-component and apolipoprotein E, and others from the tissue matrix (i.e., glycosaminoglycans), their specific roles in the polymerization-deposition process are not certain. They may be intrinsically involved in deposition or more likely, at least for P-component and Apo E, are deposited after the fact and perhaps stabilize the fibril.

also unclear what is going on while the precursors are in the systemic circulation, whether they are modified or fold in a certain way, or if their interactions with other molecules allow them to maintain their solubility. Although many of the precursors are apoproteins, or components of complex protein structures, it is unusual to find their ligands or their partner proteins in the deposits, although other proteins are co-deposited uniformly (e.g., amyloid P-component and apolipoprotein E), as are a set of glycosaminoglycans. It may be that the precursors' normal functions as apoproteins allow them to remain soluble in the circulation at relatively high concentrations.

The other feature about which we know little is why these molecules deposit and what determines the site of deposition. Fibrils are not randomly deposited throughout the body. There seems to be a specificity and a hierarchy of deposition, with each precursor having its own pattern. However, some tissues appear to be targets of several forms of fibril deposition (e.g., many precursors bind to peripheral nerve, several bind to heart, almost all deposit within vascular walls, some seem to be preferentially deposited in the kidneys). The ligand for this interaction between the amyloid precursor and the tissue substance is not clear. For some years glycosaminoglycans (GAGs), heparan sulfate containing molecules in particular, have been suspected of having an important role in deposition (7); in fact, Virchow's original iodine-staining technique probably recognized this, rather than the fibrillar component of the deposits. Recent data in experimentally induced murine AA amyloid have indicated that small molecular weight GAG analogs appear to reduce the rate and extent of deposition (8).

A number of precursors show increased synthesis. In the course of the acute phase response in inflammation every one synthesizes and secretes a large amount of SAA, but only a small proportion of chronically stimulated individuals develop amyloid. Are there other, possibly genetically controlled, factors involved? The beta protein of Alzheimer's disease is encoded on chromosome 21 and individuals who are trisomic for 21 universally display cerebral amyloid and Alzheimer's changes after the age of 45 years. Hence a gene dose effect is clearly evident in some amyloids. When clones producing immunoglobulin light chains with certain primary structures expand, the increase in the amyloidogenic clonal product is deposited in tissues and compromises organ function. In some local forms of amyloid deposition (e.g., medullary carcinoma of the thyroid) it is clear that there is local overproduction, or at least synthesis in excess of the local capacity to either catabolize or transport the amyloidogenic precursor.

Only one form of amyloid has been associated with decreased systemic excretion, the beta-2 microglobulin deposits associated with chronic dialysis. With increasing duration of dialysis, the frequency of amyloid deposition increases (9). The deposited protein has a normal amino acid sequence. Polymers of the protein precipitate in several tissues, producing clinical carpal tunnel syndrome and osteonecrosis. This is clearly an iatrogenic disease that did not appear to exist prior to the development of dialysis technology. However, there now have been one or two cases reported in which patients with chronic renal disease have accumulated sufficient amounts of the protein to produce tissue amyloid deposits without having undergone any form of dialysis.

The amyloids, which will probably provide the most information concerning the elements required for proteins to remain soluble underphysiologic conditions, are those that result from structural gene mutations that produce amino acid changes in their normal primary structures. There have been a variety of mutations in transthyretin, cystatin C, gelsolin, the beta protein of Alzheimer's disease (APP), fibrinogen, apolipoprotein A1, and lysozyme, which have been associated with autosomal dominant, familial amyloidoses (10). Some of these precursors (e.g., transthyretin and APP) are associated with both mutant and normal structure amyloids.

All of the amyloid precursors share some general features. They are all relatively small. The largest one is the prion protein deposited in Creutzfeldt-Jakob disease and scrapie, which has a molecular weight of approximately 30,000 (11). Occasionally some of light chain fibril subunits will also be that large, but the other protein precursors are relatively small, in the 5,000 to 15,000 dalton range. They all have a fair amount of beta-pleated sheet structure, but not all proteins with a lot of beta-pleated sheet structure are amyloidogenic. Thus, beta sheets are important but not sufficient. In fact, the normal prion protein (PrpC), in the absence of the mutant protein, has little beta sheet structure. When it folds to make the fibril it becomes a highly beta-pleated structure even though it has the same amino acid sequence that it had in the extended form. This phenomenon emphasizes the fact that the amyloidoses are disorders of secondary protein structure, and that understanding them will give us greater insight into the relationship between protein structure and physiologic function.

Each of the amyloidogenic precursors has been isolated. In most instances the purified protein can be incubated in a test tube under certain conditions, and it will form Congo Red–binding, ultrastructurally defined amyloid fibrils. In many cases the proteins have been cleaved enzymatically to yield so-called core peptides, which have all the fibrillogenic capacity of the intact protein. The amino acid sequences of the "core peptides" derived from different precursors are neither similar nor identical. They vary in size and they have no homology at the amino acid level.

Several autopsy studies have shown that the frequency of amyloid deposition in tissues and the number of tissues affected increase with increasing age (12). However, these age-associated amyloids are not all derived from the same precursor. The amyloid in the pancreas is islet amyloid protein precursor (or amylin); the amyloid in the cardiac atria of the heart is related to atrial natriuretic factor and that in the ventricles is related to transthyretin. The amyloid in the aorta is related to apolipoprotein A1, and in the brain it is APP. You can infer from this either that the same process is involved in all forms of amyloidogenesis, independent of the nature of the precursor, or that there is a separate amyloid pathway for each precursor and the various mechanisms age at similar, but independent, rates.

Transthyretin: A Precursor In Acquired and Hereditary Disease

Transthyretin, previously called thyroxin-binding prealbumin, is a serum transporter of thyroxine and retinol-charged, retinol-binding protein in the serum. It is synthesized in the liver, the choroid plexus, and the eye. In contrast to the SAA protein (the AA amyloid precursor), which increases in inflammation as a positive acute-phase reactant, transthyretin behaves as a negative acute-phase reactant, decreasing during inflammation and with chronic malnutrition. In the serum it circulates as a tetramer of identical subunits of 127 amino acids. It is encoded by a single gene on chromosome 18. The complete amino acid sequence and the gene sequence are known. It was one of the first molecules crystallized

(Fig 2). It has eight strands that are arranged in two sheets. There is a little alpha helix and a lot of random coil between the beta sheets. The molecule is associated with two major forms of amyloid. One is a familial, autosomal dominant set of polyneuropathies and cardiomyopathies, originally described in Portugal, but now known to have a worldwide distribution (10). These usually appear in individuals in the fourth decade and later. Over the past 15 years more than 40 mutations in transthyretin have been described. They are associated with either neuropathy, cardiomyopathy, or both. You would think in a molecule of 127 amino acids where you can identify 40 mutations, each producing physiologic instability, that we would understand the mechanism behind the process. We do not.

The amyloidogenic mutations are spread throughout the molecule, and therefore it does not seem that a single physical-chemical change can be produced by all of these. The frequency of mutations in amino acids 45 to 65 is greater than that in the rest of the molecule. The A strand and the H strand have the lowest density of amyloidogenic changes. It is possible that the paucity of mutations in one or both of those regions reflects the fact that they may be necessary for transthyretin fibrillogenesis. Consistent with this notion is the observation that the A and H strands have core peptide properties. However, it is also true that the A and H strands are involved in dimer-dimer interactions; hence, mutations in those regions may not be tolerated with respect to function.

It has been suggested that because the precursor molecules are synthesized in a soluble form that the defect must be at the time of denaturation before normal catabolic degradation and that the mutations render these molecules unstable in the unfolded state and thus subject to fibrillogenic aggregation, so-called "off-pathway folding." It may be the unfolding of the CD loop, which encompasses a small region of beta protein (the C and D beta strands) plus this loop, gets it out of the way, and allows the core peptides on adjacent molecules to interact.

In addition to these relatively rare familial autosomal dominant diseases, transthyretin is also the fibril precursor in what used to be called senile cardiac amyloidosis, generally considered to be a sporadic (i.e., nongenetically determined) incidental autopsy finding of no clinical significance, with a high prevalence over the age of 80. Ten years ago it became clear that the protein was also deposited in the lungs, in blood vessels, and occasionally elsewhere; the name was changed from senile cardiac to senile systemic amyloid (SSA) (13). The deposition of the same protein in similar tissues, in disorders that were both genetically determined and apparently acquired, drew our interest. It became even more interesting when fibrils, which were isolated from the heart of an individual with apparently sporadic senile systemic amyloid, exhibited an amino acid substitution of isoleucine for valine at transthyretin position 122 (14). The fact that the amino acid change was in the extracted fibril suggested that it was significant.

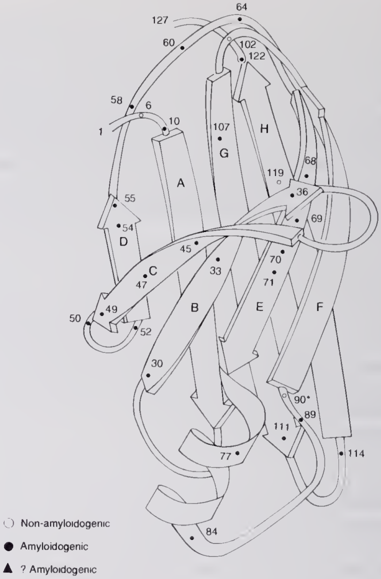


Fig. 2. A schematic drawing of the crystal structure of the transthyretin monomer with the sites of the first 29 amino acid substitutions reported. The closed circles are amyloidogenic mutations whereas the open circles are not. The total number of amyloidogenic substitutions is now more than 50.

The Molecular Genetics of SSA

The normal transthyretin gene has a digestion site for the enzyme *Mae*III, which encompasses the codon for valine 122 (15). The substitution of isoleucine for valine at that position required a disruption of the *Mae*III site, which could be analyzed by either restriction digestion and Southern blotting of genomic DNA or amplification of transthyretin exon 4 followed by restriction digestion of the amplified product. Normal transthyretin will be digested with the enzyme; the mutant transthyretin will not.

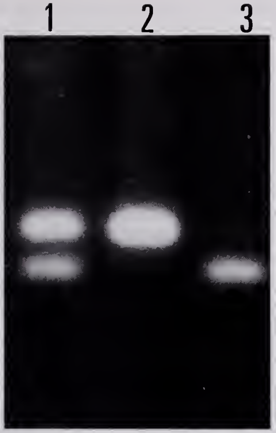


Fig. 3. The results of the *Fok*I, PCR-PIRA assay for transthyretin Ile-122 are shown in this ethidium bromide-stained agarose mini gel. Lane 1 contains amplified exon 4 from a heterozygote; lane 2 contains a homozygous normal; lane 3 contains a homozygous Ile-122 individual. (The figure was supplied by Dr. Daniel R. Jacobson.)

DNA, from the patient in whom the amino acid substitution was demonstrated, either in the genomic form or as the amplified fourth exon, showed no digestion, indicating that no copies of the normal allele were present and that the valine-isoleucine substitution, which resulted from a G-to-A transition removing the *Mae*III restriction site, was present on both transthyretin alleles. The observation suggested that perhaps the patients with SSA, in contrast to the other transthyretin disorders that were inherited in an autosomal dominant fashion, displayed recessive inheritance, with two copies of the mutant gene and no copies of the normal gene. The apparently sporadic nature of SSA would be accounted for by the requirement for the contribution of one mutant allele from each of the heterozygous parents, a relatively rare event, and the absence of a clear family history. Subsequently, two other laboratories made similar observations. Surprisingly, the three individuals initially reported from the three different laboratories were all African American, a striking observation because hereditary amyloid had never been reported in individuals of African descent before these reports.

To define the frequency of the transthyretin Ile-122 allele in the African American population we developed a polymerase chain reaction assay diagnostic for the mutation. Exon 4 was amplified using a 3′-primer that incorporated a mismatched base chosen to introduce a digestion site for the enzyme *Fok*I into amplified DNA containing the Ile allele (16). In contrast to the original *Mae*III assay, in which the normal allele is cut and the mutant allele remains intact, with *Fok*I the mutant allele is digested while the wild type (normal) gene remains intact, changing the assay to a positive rather than negative readout that is less sensitive to confounding by nonspecific enzyme inhibition (Fig. 3).

In 177 DNAs, gathered from a variety of sources, we found four heterozygotes, a frequency of 2.2%, an allele frequency of 0.011 (95% confidence interval 0.003–0.035). Subsequent studies of a much larger population sample indicated a somewhat higher frequency (Buxbaum, unpublished).

What is the clinical and pathologic significance of carrying one or two doses of the transthyretin Ile-122 allele? A report from the Los Angeles County–U.S.C. Medical Center detailed the findings in autopsies of 53,000 individuals carried out from 1949 to 1952, emphasizing the ethnic distribution of all forms of amyloid (17). The subjects were classified phenotypically as Caucasian or Black and as Hispanic by having Hispanic surnames. (The latter population in Los Angeles is primarily Mexican.) The frequency of primary or secondary amyloid in the Black population was lower than in either Caucasians or Hispanics.

There were no cases of cardiac amyloid of this type detected at autopsy before the age of 60 years. More to the point was the observation of our studies that, of those cases categorized as senile cardiac (i.e., amyloid deposits restricted to the heart in individuals in the older age group), the frequency in the African American population was twice what it was in the Caucasian and virtually 10 times that in the Hispanic population. In collaboration with Dr. Buck we found that the percentage of Caucasian individuals with senile systemic (cardiac) amyloid increases with increasing age. The same is true in African Americans except that the prevalence is higher in every decade. It appears that if you shift the Caucasian incidence curves two decades earlier, you can almost superimpose the prevalence plots of the African Americans and Caucasians.

What is the clinical significance of the findings? There is a substantial literature, which has emerged over the past 20 years, documenting the clinical importance of cardiac amyloid deposits in the elderly. In a study from Middlesex, England, Hodkinson and Pomerance (18) showed that substantial amounts of ventricular amyloid found at autopsy are associated with significant increases in the incidence of both atrial fibrillation and congestive heart failure. In this country, in a study of the oldest old (individuals above the age of 90) from Olmstead County, Minnesota, more than 20% of the individuals autopsied had ventricular amyloid, and in that population it was the cause of death in almost 50% (19).

The limited clinical data available from the Los Angeles autopsies revealed that of the 23 patients with cardiac amyloid and heart failure, one third of them had no other anatomic heart disease. Similarly, a third of the patients with atrial fibrillation in the presence of amyloid had no other heart disease. With respect to ventricular conduction defects, again one third of the patients had no other anatomic heart disease. Hence isolated cardiac amyloid deposition in the elderly has considerable clinical import. Its role as a cofactor when it co-exists with other forms of heart disease is not yet clear.

To return to the genetic epidemiology of the disorder we must reconsider our original hypothesis that SSA in African Americans is a recessive disease. The Los Angeles data indicated that the prevalence was 10 times greater than predicted. Why? We have determined the allele frequency in the population from a sample with adequate power and an assay of exquisite specificity. We have confirmed the transthyretin nature of the cardiac deposits in the Los Angeles autopsies. Hence, Dr. Buck's criteria for anatomic diagnosis were 95% correct in categorizing these patients as senile cardiac disease. Examination of DNA extracted from archival material provided by Dr. Buck indicate that transthyretin Ile-122 is significantly more common among autopsied individuals with SSA than it is in among those of similar age and ethnic background whose autopsies did not reveal any cardiac deposition. It is also clear, however, that not all late-onset cardiac amyloid deposition in the population could be accounted for by the presence of the transthyretin Ile-122 allele.

Conclusions

Approximately 4% of African Americans are heterozygous for an amino acid substitution in which isoleucine is present in the serum protein transthyretin. Cardiac amyloid deposition after age 60 is more common in African Americans than in other ethnic groups. The precursor is transthyretin. Much of the late-onset cardiac amyloid deposition is related to the presence of the Ile-122 allele. The remainder represent normal sequence or other mutations that have not yet been identified. The mutation represents an absolute risk for anatomic amyloid deposition in the heart and to a certain extent, clinical heart disease, yet does not account for the majority of instances of late-onset cardiac amyloid deposition. It remains to be determined whether those individuals bearing the mutation have more severe disease than those with deposits containing only normal sequence transthyretin. Conceptually, amyloid related to transthyretin Ile-122 should be considered another form of familial amyloidotic cardiomyopathy, whereas normal sequence cardiac deposition is better designated SSA. Clinically, however, at the time of presentation, in the absence of a positive family history, it is difficult to distinguish between the two. Although endomyocardial biopsy with antibody staining provides the "gold standard" for pathologic diagnosis, molecular genetic analysis is required for a definitive etiologic diagnosis.

Questions and Answers

Q: Could you touch upon the coagulopathy associated with primary amyloid?

A: There are several types of coagulopathies that are associated with amyloid. The most common reason for bleeding is the vascular fragility that is a direct function of amyloid deposition in the vessels. In disorders in which there is neuropathy with a resultant decrease in tissue tone and vascular support, the effect of the vascular fragility of the vessels is accentuated. The true coagulopathies are associated with the immunoglobulin type of amyloid (AL). A number of instances of factor X deficiency have been reported. It has been hypothesized that the responsible mechanism is binding of factor X to the amyloid fibrils, because intravenously administered labeled factor X seemed to localize to organs with substantial amyloid deposits. However, in other patients, in vitro studies failed to demonstrate factor X binding to fibrils isolated from the spleen of a patient who exhibited factor X deficiency in vivo. It may be that in the latter studies the purification of the fibril resulted in the reduction or elimination of the binding capacity for factor X noted in the in vivo studies.

Q: There are many reports dealing with reversal of amyloid, and the prevention of renal amyloid with colchicine.

A: Present therapy for light chain amyloid is identical with that for multiple myeloma. Patients are treated with melphalan and prednisone in a standard myeloma regimen, with doses calculated on the basis of body mass and cycles every 5 weeks. Studies, thus far presented only in abstract, show that, after stratifying for organ involvement, these regimens result in longer survival. I think as myeloma therapies become more aggressive, and if the more aggressive therapies prove to be effective, amyloid patients will get treated in the same way. This is reflected in a recent report of a patient with AL amyloid being treated with intensive chemotherapy followed by stem cell replacement. Colchicine offers no additional benefit in light chain (AL) amyloid. The use of colchicine is restricted to AA amyloid, and in particular, familial Mediterranean fever, where it has been extremely effective in both aborting the febrile episodes and preventing renal amyloid deposition, especially when administered prior to the development of amyloid. It has not been effective in any other form of amyloid. It is not clear if it has any effect on the development of AA in any other setting.

Q: How effective are subcutaneous and rectal biopsy for diagnosing all forms of amyloid?

A: In the systemic forms of amyloid, you get 90% yield using a combination of both, assuming that you have good samples including vessels in the rectal biopsy and that you have laboratory personnel who know who to process the material. In addition to getting Congo Red positivity, if it is done in a laboratory that has antibodies to all the precursors, you can not only determine that amyloid is present, you can also tell what kind of amyloid.

Q: Can you cause the mutated form of transthyretin to solubilize?

A: We have not done that, but other investigators are using the process of in vitro fibrillogenesis to examine approaches to either the prevention of fibril formation or fibril solubilization.

Q: Are there some conditions that would allow transthyretin to solubilize?

A: In vitro fibrillogenesis studies are carried out under conditions that are highly unphysiologic. They are usually done at high protein concentrations, which are never achieved in vivo, and at pH values that which may only be achievable inside lysosomes. There is clinical evidence that abrogation of precursor synthesis will ultimately result in solubilization of tissue fibrils.

Q: How can you be sure of the genetic makeup of the Hispanic population in Los Angeles?

A: That is more of an issue in this part of the country than it is in California. The Hispanic population in Los Angeles is largely Mexican.

References

1. Buxbaum JN. The Amyloidoses. In: Schumacher HR Jr, Klippel JH, Koopman WJ, eds. Primer on the rheumatic diseases. 10th ed. Atlanta: Arthritis Foundation, 1993: 234.
2. Pras M, Schubert M, Zucker-Franklin D et al. The characterization of soluble amyloid prepared in water. J Clin Invest 1968; 47:924–933.
3. Westermark P, Stenkvist B, Natvig JB et al. Demonstration of protein AA in subcutaneous fat tissue obtained by fine needle biopsy. Ann Rheum Dis 1979; 38:68–71.
4. Arapakis G, Tribe CR. Amyloidosis in R.A. investigated by means of rectal biopsy. Ann Rheum Dis 1963; 22: 256–262.
5. Mutru O, Laakso M, Isomaki H, Koota K. Ten year mortality and causes of death in patients with rheumatoid arthritis. Br Med J 1985; 290:1797–1799.
6. Kelly JW, Lansbury PT Jr. A chemical approach to elucidate the mechanism of transthyretin and beta-protein amyloid fibril formation. Int J Exp Clin Invest 1994; 1:186–205.
7. Kisilevsky R. Proteoglycans, glycosaminoglycans, amyloid-enhancing factor, and amyloid deposition. J Intern Med 1992; 232:515–516.
8. Kisilevsky R, Lemieux LJ, Fraser PE et al. Arresting amyloidosis in vivo using small molecule anionic sulphonates or sulphates: implications for Alzheimer's disease. Nat Med 1995; 1:143–148.
9. Gejyo F, Yamada T, Odani S et al. A new form of amyloid protein associated with chronic hemodialysis was identified as B_2-microglobulin. Biochem Biophys Res Commun 1985; 129:701–706.
10. Jacobson DR, Buxbaum JN. Genetic aspects of amyloidosis. Adv Hum Genet 1991; 20:69–123.
11. Prusiner SB. Molecular biology of prion diseases. Science 1991; 252:1515–1522.
12. Wright MD, Calkins E, Breen W et al. Relationship of amyloid to aging. Review of the literature and systematic study of 83 patients derived from a general hospital population. Medicine 1969; 48:39–60.
13. Pitkänen P, Westermark P, Cornwell GG III. Senile systemic amyloidosis. Am J Pathol 1984; 117:391–399.
14. Gorevic PD, Prelli FC, Wright J et al. Systemic senile amyloidosis. Identification of a new prealbumin (transthyretin) variant in cardiac tissue: immunologic and biochemical similarity to one form of familial amyloidotic polyneuropathy. J Clin Invest 1989; 83:836–843.
15. Jacobson DR, Gorevic PD, Buxbaum JN. A homozygous transthyretin variant associated with senile systemic amyloidosis: evidence for a late-onset disease of genetic etiology. Am J Hum Genet 1990; 47:127–136.
16. Jacobson DR. A specific test for transthyretin 122 (Val-Ile) based on PCR-primer introduced restriction analysis (PCR-PIRA): confirmation of the gene frequency in Blacks. Am J Hum Genet 1992; 50:195–198.
17. Buck FS, Koss MN, Sherrod AE et al. Ethnic distribution of amyloidosis: an autopsy study. Mod Pathol 1989; 2: 372–377.
18. Hodkinson HM, Pomerance A. The clinical significance of senile cardiac amyloidosis: a prospective clinico-pathological study. Q J Med 1977; XLVI:381–387.
19. Lie JT, Hammond H. Pathology of the senescent heart: anatomic observations on 237 autopsy studies of patients 90 to 105 years old. Mayo Clin Proc 1988; 63:552–564.

The Art and Science of Consultation

Consultation in Sleep Medicine

Lee K. Brown, M.D.

The role of the consultant in sleep medicine is undoubtedly in transition, paralleling changes taking place in specialty medicine overall. The change involves a shift in responsibilities, with the generalist ("gatekeeper") assuming much of the burden for evaluation and treatment of disorders that were formerly the purview of the specialist. In this paradigm, specialty referral is reserved for particularly complex or perplexing problems or when a specialty procedure is required. The paradigm clearly has its seamy side: In many locales around the United States, "alternative" forms of medical reimbursement such as capitated care are now more nearly the norm; sending a patient for specialty evaluation and treatment carries a financial penalty for the referring physician, and such referrals become less frequent (1). However, in the case of sleep disorders medicine, some believe this paradigm shift may be long overdue (2). Sleep symptoms are ubiquitous within the general population: 38% of respondents in one large survey had a current sleep complaint, including 32% suffering from insomnia and 4% troubled by excessive somnolence (3). At the same time, the personal and societal consequences of not addressing sleep health can be monumental (4).

Clearly, many aspects of sleep medicine should be practiced at the primary care level. Unfortunately, there is information to suggest that the physician in general practice has a woeful track record in addressing these symptoms. In one study of a primary care clinic in the state of Washington, Dement and colleagues (5) reviewed the complete charts of 1000 patients under the care of 20 different general practice physicians during 1991; they were unable to identify a single patient who received a specific sleep disorders diagnosis.

There are probably several reasons for this failure to address sleep problems in general practice. It would appear that less and less time is being allowed for patient encounters, and the primary care physician is being exhorted to include more and more services during that time. In addition, very little training has traditionally been provided to medical students concerning sleep and sleep disorders. A recent survey of U.S. medical schools revealed that less than 2 hours of instruction, on average, was devoted to sleep medicine in 4 years of medical education (6). Schaffner (7), in a previous article in this series, stated that one of the most important goals of a consultation is to teach some aspect of the specialty to the referring physician. Similarly, this article will outline the elements of a sleep medicine consultation with the expectation that some aspects will be incorporated into general practice, whereas others will highlight the need for appropriate specialty referral.

A Whirlwind Tour of Sleep Medicine

The wide scope of sleep disorders medicine perhaps can be appreciated best by reviewing the latest nosology of sleep medicine (**Tables 1 to 3**) as embodied in the *International Classification of Sleep Disorders* published in 1990 (8). Three major categories of sleep disorders are recognized: *Dyssomnias* cause either of the two principal sleep-related symptoms (insomnia or excessive sleepiness), and are not the result of another medical or psychiatric illness. *Parasomnias* consist of behaviors or physical symptoms that are associated with sleep. They generally do not result in

---

From the Sleep Disorders Center, Department of Internal Medicine, St. Joseph's Hospital and Medical Center, Phoenix, Arizona, and the Division of Pulmonary and Critical Care Medicine, Mount Sinai School of Medicine, New York, New York. Address reprint requests to Lee K. Brown, M.D., Sleep Disorders Center, Department of Internal Medicine, St. Joseph's Hospital and Medical Center, 350 West Thomas Road, Phoenix, AZ 85013.

TABLE 1

Outline, International Classification of Sleep Disorders: Dyssomnias

- A. *Intrinsic sleep disorders*
    - Psychophysiologic insomnia
    - Sleep state misperception
    - Idiopathic insomnia
    - Narcolepsy
    - Recurrent hypersomnia
    - Idiopathic hypersomnia
    - Post-traumatic hypersomnia
    - Obstructive sleep apnea syndrome
    - Central sleep apnea syndrome
    - Central alveolar hypoventilation syndrome
    - Periodic limb movement disorder
    - Restless legs syndrome
    - Intrinsic sleep disorder NOS
- B. *Extrinsic sleep disorders*
    - Inadequate sleep hygiene
    - Environmental sleep disorder
    - Altitude insomnia
    - Adjustment sleep disorder
    - Insufficient sleep syndrome
    - Limit-setting sleep disorder
    - Sleep-onset association disorder
    - Food allergy insomnia
    - Nocturnal eating (drinking) syndrome
    - Hypnotic-dependent sleep disorder
    - Stimulant-dependent sleep disorder
    - Alcohol-dependent sleep disorder
    - Extrinsic sleep disorder NOS
- C. *Circadian rhythm sleep disorders*
    - Time zone change (jet lag) syndrome
    - Shift work sleep disorder
    - Irregular sleep-wake pattern
    - Delayed sleep phase syndrome
    - Advanced sleep phase syndrome
    - Non-24-hour sleep-wake disorder
    - Circadian rhythm sleep disorder NOS

Adapted from: Diagnostic Classification Steering Committee, Thorpy MJ, Chairman. International classification of sleep disorders: Diagnostic and coding manual. Rochester, MN: American Sleep Disorders Association, 1990, with permission.

TABLE 2

Outline, International Classification of Sleep Disorders: Parasomnias

- A. *Arousal disorders*
    - Confusional arousals
    - Sleepwalking
    - Sleep terrors
- C. *Parasomnias usually associated with REM sleep*
    - Nightmares
    - Sleep paralysis
    - Impaired sleep-related penile erections
    - Sleep-related painful erections
    - REM sleep-related sinus arrest
    - REM sleep behavior disorder
- B. *Sleep-wake transition disorders*
    - Rhythmic movement disorder
    - Sleep starts
    - Sleep talking
    - Nocturnal leg cramps
- D. *Other parasomnias*
    - Sleep bruxism
    - Sleep enuresis
    - Sleep-related abnormal swallowing syndrome
    - Nocturnal paroxysmal dystonia
    - Sudden unexplained nocturnal death syndrome
    - Primary snoring
    - Infant sleep apnea
    - Congenital central hypoventilation syndrome
    - Sudden infant death syndrome
    - Benign neonatal sleep myoclonus
    - Other parasomnia NOS

Adapted from: Diagnostic Classification Steering Committee, Thorpy MJ, Chairman. International classification of sleep disorders: Diagnostic and coding manual. Rochester, MN: American Sleep Disorders Association, 1990, with permission.

the cardinal symptoms that define the dyssomnias, but their manifestations can be quite troublesome to either the patient or observers. *Sleep disorders associated with medical/psychiatric disorders* encompass the physical or emotional illnesses that can interfere with sleep or wakefulness and therefore cause insomnia or excessive somnolence. A possible fourth category, *proposed sleep disorders,* consists of disorders that have been described anecdotally but not investigated to the extent that their existence can be firmly established. Generally, these disorders (assuming they exist at all) are rare; consequently this category is not represented in the tables and will not be further discussed.

Dyssomnias

This category is further subdivided into *intrinsic sleep disorders* that arise primarily from processes within the patient, *extrinsic sleep disorders* that result mainly from outside influences,

TABLE 3

Outline, International Classification of Sleep Disorders: Sleep Disorders Associated with Medical/Psychiatric Disorders

A. *Associated with mental disorders*
- Psychoses
- Mood disorders
- Anxiety disorders
- Panic disorder
- Alcoholism

B. *Associated with neurologic disorders*
- Cerebral degenerative disorders
- Dementia
- Parkinsonism
- Fatal familial insomnia
- Sleep-related epilepsy
- Electrical status epilepticus of sleep
- Sleep-related headaches

C. *Associated with other medical disorders*
- Sleeping sickness
- Nocturnal cardiac ischemia
- Chronic obstructive pulmonary disease
- Sleep-related asthma
- Sleep-related gastroesophageal reflux
- Peptic ulcer disease
- Fibrositis syndrome

Adapted from: Diagnostic Classification Steering Committee, Thorpy MJ, Chairman. International classification of sleep disorders: Diagnostic and coding manual. Rochester, MN: American Sleep Disorders Association, 1990, with permission.

and *circadian rhythm sleep disorders* that arise from a lack of synchrony between the patient's sleep/wake pattern and that dictated by the surrounding community.

Intrinsic sleep disorders may stem from either behavioral or physical processes. *Psychophysiologic insomnia* is a behavioral disorder with two components: anxiety about life stressors that manifests as hyperalertness and a learned association between the sleeping environment and wakefulness. The latter association may relate to the physical environment, that is the bed, bedroom, or presleep rituals; or may be internal in nature: the harder one tries to fall asleep, the less likely one will be able to accomplish this task given the "performance anxiety" that this engenders. With the exception of sleep disturbance associated with psychiatric illness, psychophysiologic insomnia is the most common cause of that symptom in patients presenting to the sleep laboratory for evaluation (9). *Idiopathic insomnia* is a chronic disorder of initiating and maintaining sleep, typically beginning in childhood. The pathophysiology of idiopathic insomnia is felt to involve a defect in the sleep-generating system of the central nervous system, and some familial tendencies have been demonstrated. Patients with *sleep state misperception* complain of insomnia, but laboratory testing reveals relatively normal sleep architecture. Nonetheless, the patient will maintain that sleep did not occur during the recording night. The pathophysiology is not known, but may involve subtle abnormalities in the sleep electroencephalogram that are currently undetectable, or may be a form of hypochondriasis.

Nacrolepsy is a neurologic disease characterized by a defect in the control of sleep-wake transitions. This defect involves not only the boundary between wakefulness and non–rapid eye movement (REM) sleep, but also between wakefulness and REM (dreaming) sleep. Patients with narcolepsy are abnormally sleepy when they wish to remain awake, but paradoxically may have disturbed nocturnal sleep. In addition, they are subject to the REM-intrusive phenomena of cataplexy, sleep paralysis, and hypnagogic hallucinations. In cataplexy and sleep paralysis, the muscle atonia usually confined to REM sleep intrudes upon wakefulness. This occurs in response to strong emotion (cataplexy) or shortly before and after a sleep period (sleep paralysis). Hypnagogic hallucinations are vivid dreams associated with sleep onset and represent the dreaming component of REM sleep intruding upon wakefulness. The etiology of narcolepsy is not known; however, there is a strong association with certain HLA types, which suggests an autoimmune pathogenesis.

Patients with *idiopathic hypersomnia* exhibit the excessive daytime sleepiness generally associated with narcolepsy, but REM-intrusive phenomena are not observed, and nocturnal sleep is essentially normal. As in narcolepsy, it is postulated that these patients possess a defect in the central nervous system circuits responsible for the control of sleep and wakefulness. *Post-traumatic hypersomnia* is temporally associated with central nervous system trauma and presumably results from injury to areas involved in the control of sleep and wakefulness. Slow resolution usually occurs over the course of months following the injury. *Recurrent hypersomnia,* also known as *Kleine-Levin syndrome,* is a rare disease generally involving adolescent males. Periods of severe hypersomnia (the patient may sleep 18 hours per day or more) lasting days to weeks are separated by months of remission. The disorder may be accompanied by hypersexuality and hyperphagia (the classic Kleine-Levin syndrome), or the hypersomnolence may occur in isolation.

Obstructive sleep apnea syndrome is an important cause of dyssomnic symptoms, predominantly excessive daytime sleepiness. Recent data point to a prevalence of 2% to 4% in the working, adult population (10). Periodic obstruction to airflow in the oropharyngeal and/or hypopharyngeal airway leads to recurrent, brief arousals during sleep, as well as nocturnal hypoxemia. Although the arousals are generally too brief to be recalled by the patient the next day, an occasional patient will experience longer-lasting periods of wakefulness and complain of sleep-maintenance insomnia. In addition to the dyssomnic symptoms, patients with significant obstructive sleep apnea have a demonstrated increase in mortality (11), which may stem from a higher incidence of myocardial infarction and stroke (12), or automobile and workplace accidents (13). Patients with *central sleep apnea syndrome* exhibit periodic cessations in ventilatory effort, which may lead to arousals from sleep. This disorder occurs much less frequently than obstructive sleep apnea, and for unknown reasons is more commonly associated with complaints of insomnia, restless sleep, and choking during sleep. Daytime somnolence may also occur. *Central alveolar hypoventilation,* also poetically termed *Ondine's curse,* may occur in patients with or without known neurologic disease. A variety of neuromuscular disorders can impair central control of ventilation or respiratory muscle strength, whereas in patients without diagnosed neurologic illness, an occult neurologic disorder is assumed. Long periods of hypoventilation and hypoxemia are present during sleep, resulting in frequent arousals that may be perceived by the patient (insomnia) or be manifest only as daytime somnolence.

Patients with *periodic limb movement disorder* demonstrate stereotyped movements of the legs (or occasionally, arms) that occur in a regular pattern and frequently result in arousal from sleep. As in obstructive sleep apnea syndrome, the arousals may be so brief that their only manifestation is daytime somnolence, or they may be sufficiently prolonged to generate complaints of insomnia. The incidence increases with age; in one study, 34% of randomly selected subjects aged 65 or older had the disorder (14). The etiology is unknown, although a neuropathy of some type is suspected due to the increased occurrence of this disorder in association with uremia and other metabolic disorders known to cause peripheral neuropathies. The hallmark of *restless legs syndrome* is the complaint of an irresistable urge to move one's legs (and occasionally, the arms), often associated with a poorly described discomfort that some patients characterize as paresthesias, "aching," or "crawling." The disorder commonly results in sleep onset or maintenance insomnia and, depending on the amount of sleep loss, may be accompanied by daytime somnolence as well. Interestingly, virtually all patients with restless legs syndrome will demonstrate periodic limb movements during sleep, and the disorder is associated with the same systemic illnesses, again suggesting a neuropathic etiology.

Extrinsic sleep disorders all relate to behaviors practiced by the patient or those in his environment that result in disturbed sleep. *Inadequate sleep hygiene* refers to insomnia resulting from a multitude of diurnal and bedtime behaviors that interfere with sleep (Table 4). These include excessive daytime napping such that sleepiness at bedtime is inadequate to allow sleep; irregular sleep-wake schedules that interfere with the normal circadian cycling of sleepiness; tobacco, alcohol, and caffeine use prior to bedtime; and practices that promote arousal in bed rather than sleepiness, such as a noisy environment (e.g., keeping the television on) or performing job-related activities in bed. *Environmental sleep disorder* is invoked when an environmental phenomenon is responsible for sleep disturbance (e.g., noise, uncomfortable temperature, lack of security) and when more global disregard of sleep hygiene is not present. *Altitude insomnia* is also essentially an environmental disorder; headache, breathing disturbances, and other physical complaints related to the hypoxia of an acute move to high altitude result in interrupted sleep and insomnia. *Insufficient sleep syndrome* results when another precept of sleep hygiene is violated: the patient does not allow enough time in bed to achieve adequate sleep. *Adjustment sleep disorder* is diagnosed when a discrete life event (e.g., a new job or a test at school) engenders anxiety and hyperarousal at bedtime.

Limit-setting sleep disorder, sleep-onset association disorder, food allergy insomnia, and *nocturnal eating* (*drinking*) *syndrome* are essentially pediatric sleep disorders. In limit-setting sleep

TABLE 4
Sleep Hygiene Instructions

- Homeostatic drive for sleep
  - Avoid naps, except for a brief 10- to 15-min nap 8 h after arising; but check with your physician first, because in some sleep disorders naps can be beneficial.
  - Restrict sleep period to average number of hours you have actually slept per night in the preceding week. Quality of sleep is important. Too much time in bed can decrease quality on subsequent night.
  - Get regular exercise each day, preferably 40 min each day of an activity that causes sweating. It is best to finish exercise at least 6 h before bedtime.
  - Take a hot bath to raise your temperature 2°C for 30 min within 2 h before bedtime. A hot drink may help you relax as well as warm you.
- Circadian factors
  - Keep a regular time out of bed 7 days a week.
  - Do not expose yourself to bright light if you have to get up at night.
  - Get at least one half hour of sunlight within 30 min of your out-of-bed time.
- Drug effects
  - Do not smoke to get yourself back to sleep.
  - Do not smoke after 7 p.m., or give up smoking entirely.
  - Avoid caffeine entirely for a 4-week trial period; limit caffeine use to no more than three cups no later than 10 a.m.
  - Light to moderate use of alcoholic beverages. Alcohol can fragment sleep over second half of sleep period.
- Arousal in sleep setting
  - Keep clock face turned away, and do not find out what time it is when you wake up at night.
  - Avoid strenuous exercise after 6 p.m.
  - Do not eat or drink heavily for 3 h before bedtime. A light bedtime snack may help.
  - If you have trouble with regurgitation, be especially careful to avoid heavy meals and spices in the evening. Do not retire too hungry or too full. Head of bed may need to be raised.
  - Keep your room dark, quiet, well ventilated, and at a comfortable temperature throughout the night. Ear plugs and eye shades are OK.
  - Use a bedtime ritual. Reading before lights-out may be helpful if it is not occupationally related.
  - List problems and one-sentence next steps for the following day. Set aside a worry time. Forgive yourself and others.
  - Learn simple self-hypnosis to use if you wake up at night. Do not try too hard to sleep; instead, concentrate on the pleasant feeling of relaxation.
  - Use stress management in the daytime.
  - Avoid unfamiliar sleep environments.
  - Be sure mattress is not too soft or too firm, pillow is right height and firmness.
  - An occasional sleeping pill is probably all right.
  - Use bedfoom for sleep; do not work or do other activities that lead to prolonged arousal.
  - If possible, make arrangements for care-giving activities (children, others, pets) to be assumed by someone else.

From: Zarcone VP Jr. Sleep hygiene. In: Kryger MH, Roth T, Dement WC, eds. Principles and Practice of Sleep Medicine. 2nd ed. Philadelphia: W. B. Saunders, 1994:545, with permission.

disorder, the child attempts to delay bedtime by repeatedly requesting attention by the parent or caretaker. This takes the form of excessive and inappropriate requests for a bedtime story, drink of water, trip to the bathroom, etc. Treatment requires the firm refusal of the parent to give in to the demands. Children with sleep-onset association disorder have learned to associate sleep onset with a particular set of circumstances, and will not fall asleep unless these circumstances are met. Examples include falling asleep only in a parent's bed, requiring a bedtime bottle, or being rocked by the parent or caretaker. In food allergy insomnia, sensitivity to food allergens (usually proteins in cow's milk), result in physical symptoms that disturb sleep. Children with nocturnal eating (drinking) syndrome awaken repeatedly, presumably by hunger, and require a bottle or feeding to fall back asleep. The disorder is considered to be the result of a learned behavior, and may also occur in adults.

Hypnotic-, stimulant-, and alcohol-dependent sleep disorders are syndromes associated with the use of the respective agents at bedtime. Regular hypnotic use frequently engenders tolerance and results in dosage escalation in order to produce sleep. Worsening insomnia and anxiety ensue when withdrawal occurs as the drug is metabolized, either during the day (in the case of agents with long half-lives) or even later that same night (in short-acting agents). A similar course of events takes place when alcohol is abused to induce sleep; nocturnal awakening with sleep-maintenance insomnia is common as the ethanol is metabolized during the early part of the night and withdrawal takes place. Stimulant medication, both socially acceptable (caffeine) and illicit (cocaine, amphetamine), results in hyperarousal when taken adjacent to bedtime and interferes with sleep onset. Withdrawal from these agents can result in hypersomnolence.

Our understanding of the circadian rhythm sleep disorders is based on the discovery that sleepiness varies in a regular, predictable manner during the course of 24 hours. The internal "clock" that appears to control this rhythm resides in the suprachiasmatic nucleus and seems to cycle on the basis of a period slightly longer than 24 hours. Consequently, it must be reset on a daily basis to stay in synchrony with the regular day/night cycle of the environment. A variety of stimuli (*zeitgebers* or "time givers") act as synchronizers, the most important being exposure to light at the proper time. The disorders *time zone change (jet lag) syndrome* and *shift work sleep disorder* occur when the internal circadian rhythm of sleepiness falls out of synchrony with the day/night cycle in the environment; this follows travel to another time zone or the adoption of a sleep/wake cycle not in conformance with the "normal" practice of sleeping at night and working during the day, respectively. In *irregular sleep-wake pattern,* the patient fails to maintain any regular circadian rhythm of sleepiness and may doze at any time during the day. The disorder is generally associated with organic brain disease, although depressed patients or individuals in a position to completely disregard *zeitgebers* may also develop these findings. *Non-24-hour sleep-wake disorder* is evident primarily in certain blind patients and results from their inability to perceive light. The loss of this important *zeitgeber* leads to a sleep/wake cycle longer in period than the day/night cycle in the environment. As a consequence of this "free-running" cycle, their internal circadian rhythm continually goes in and out of synchrony with the rest of the community. *Advanced sleep phase syndrome* and *delayed sleep phase syndrome* refer to individuals whose preferred sleep period is earlier or later, respectively, than that of the majority of their neighbors. Delayed sleep phase syndrome is commonly encountered in adolescents or young adults, who may adopt the practice of staying awake until late at night, and then sleep until midday.

Parasomnias

The parasomnias are subclassified in terms of the sleep stage with which they are associated; there is also a miscellaneous category for disorders not particularly associated with a specific stage. The *arousal disorders* are all associated with arousal out of slow-wave sleep, which occurs primarily during the first third of the night. *Confusional arousals* are periods of disorientation and confusion upon awakening from slow-wave sleep; *sleepwalking* consists of complex behaviors, including walking and other activities, in the same situation. *Sleep terrors* also occur out of slow-wave sleep, but the patients manifest signs of abject fear and autonomic arousal, usually accompanied by screaming or crying.

Sleep-wake transition disorders primarily occur during the passage from drowsy wakefulness to light sleep. *Rhythmic movement disorder* is a pediatric condition; these children practice a variety of repetitive body movements as they fall asleep, including head banging, body rolling, and body rocking. It is postulated that these movements provide a vestibular stimulation that induces relaxation and thus sleep. *Sleep starts* are the sudden, brief body movements shortly after sleep onset that most individuals have experienced. *Sleep talking* and *nocturnal leg cramps* are self-descriptive phenomena not necessarily confined to sleep-wake transitions, but may occur during other stages as well.

A variety of parasomnias are associated with REM sleep; some are quite common, whereas others are extremely rare. *Nightmares* are terrifying dreams that result in awakening. They affect a substantial minority of the pediatric population, and most adults report them on occasion. Unlike night terrors, nightmares occur out of REM sleep, involve vivid, often complex dreams, and predominate during the second half of the night when REM sleep is more common. *Sleep paralysis* has already been described in the context of narcolepsy. Most adults report an isolated episode at one time or another during their lives, usually associated with sleep deprivation. A rare familial form also exists. *REM sleep behavior disorder* is an uncommon condition, usually of older men, marked by the occurrence of violent behaviors during REM sleep. The patient usually reports a dream with violent content and the behaviors are essentially an enactment of that dream. Injury to self or bed partner are not uncommon during these episodes. Polysomnography reveals persistent, inappropriate muscle tone during REM sleep. The disorder can be idiopathic or associated with organic dementias. *Impaired sleep-related penile erections* refers to that polysomnographic finding in organic impotence, whereas *sleep-related painful erections* is a rare disorder that can result in complaints of insomnia and hypersomnolence as a consequence of repeated awakenings from REM sleep due to pain. *REM sleep-related sinus arrest* is an exceedingly rare disorder affecting young males, in which cardiac asystole during REM sleep results in arousal. Cardiac evaluation is unrevealing, and the etiology is unknown.

The category of *other parasomnias* represents a hodgepodge of nocturnal behaviors that do not seem to share attributes with each other or parasomnias in the previously described categories. *Sleep bruxism* is a common condition affecting both children and adults, involving grinding or clenching of the teeth during sleep. The activity may result in dental pathology or temporomandibular joint disease, and the noise may engender environmental sleep disorder in the bed partner. *Sleep enuresis* may be primary (continuous from infancy) or secondary (developing in a child who had previously been dry all night). The primary form appears to have a familial tendency; occasionally, obstructive sleep apnea is associated with the secondary form. Patients with *sleep-related abnormal swallowing syndrome* are generally adults who report frequent nocturnal awakenings associated with coughing and choking. Investigations reveals that they are unable to swallow accumulated oral secretions during sleep. The etiology of this rare disorder is unknown. *Nocturnal paroxysmal dystonia* consists of stereotyped dystonic or dyskinetic body movements that awaken the patient out of non-REM sleep. The episodes can be frequent enough to result in sleep disruption and complaints of insomnia. Pathogenesis is unknown; the frequent response to carbamazepine has suggested that it may represent a form of epilepsy, although electroencephalography is generally unrevealing. *Sudden unexplained nocturnal death syndrome* is a self-descriptive phenomenon recognized in Southeast Asian refugees to the United States; apparently it is known to occur in many Asian countries. Pathogenesis is unknown; survivors all have been in ventricular fibrillation when they were resuscitated by emergency personnel.

Primary snoring is snoring unaccompanied by apneas, hypopneas, or sleep disruption (except for the bed partner!). *Sudden infant death syndrome* is the well-recognized phenomenon of the unexplained, unexpected death of an apparently healthy infant, usually before 6 months of age. Some cases of sudden infant death syndrome may be related to *infant sleep apnea,* the occurrence of central or obstructive apneas in that age group. *Congenital central hypoventilation syndrome* is evident at birth, typically worsens during sleep, and is not associated with any known pulmonary or neuromuscular disease. Infants with *benign neonatal sleep myoclonus* exhibit myoclonic movements of the arms or legs during sleep. Neurologic studies are unrevealing in this rare, self-limited disorder.

Sleep Disorders Associated with Medical/Psychiatric Disorders

Sleep/wake symptoms may be associated with a number of mental, neurologic, and medical disorders. Insomnia and hypersomnia commonly occur in the *psychoses;* in the *mood disorders,* sleep-onset insomnia may occur in depression or mania, and sleep maintenance insomnia (early morning awakening) in depression; sleep-onset and sleep-maintenance insomnia frequently complicate the *anxiety disorders;* and panic attacks may awaken the patient with *panic disorder* from sleep as well as occur during wakefulness. Some of the effects on sleep of acute alcohol ingestion have already been described. The patient with chronic *alcoholism* develops sleep fragmentation during binges as well as upon withdrawal; some of the effects of chronic alcoholism on sleep architecture, such as the suppression of slow-wave sleep, can last for years after detoxification.

A variety of neurologic disorders are complicated by sleep pathology. *Cerebral degenerative disorders* and *parkinsonism* are associated with complaints of insomnia and sleep fragmentation. Treatment of parkinsonism with dopaminergic agents may improve sleep quality, although excessive doses may disrupt sleep. Nighttime delirium and agitation ("sundowning") in patients with *dementia* is a well-known phenomenon. Failure of circadian rhythm generation has been a postulated etiology, thus far without firm laboratory evidence. *Fatal familial insomnia* is an exceedingly rare disorder involving progressive autonomic failure and degenerative central nervous system lesions, leading to near-total sleeplessness and death.

Certain types of *epilepsy* characteristically worsen during sleep: complex partial seizures, focal motor seizures, and generalized tonic-clonic seizures often have this tendency. Occasionally, complex partial seizures (temporal lobe epilepsy) will be shown to underlie behavior thought to be due to a simple parasomnia. Rarely, frequent nocturnal seizures will result in a complaint of insomnia. Recent evidence suggests that sleep apnea may also exacerbate epilepsy (14). *Electrical status epilepticus of sleep* is a disorder of children characterized by continuous epileptiform electrical activity throughout non-REM sleep. There are usually no clinical manifestations, although the disorder may be complicated by other epilepsies and psychiatric or neurologic findings may supervene. Electrical status epilepticus of sleep generally resolves spontaneously.

Sleep-related headaches are diagnosed when more than 75% of the episodes occur out of sleep. The vascular headaches of classic migraine, cluster, and paroxysmal hemicrania all may predominate out of sleep; the closest association exists with the latter two disorders. Vascular headache induced by hypercapnia has also been attributed to the sleep apnea syndromes.

Certain medical illnesses frequently exacerbate during sleep: *ischemic heart disease, chronic obstructive pulmonary disease, asthma, gastroesophageal reflux,* and *peptic ulcer disease* are common offenders. As a result, these disorders are often associated with sleep fragmentation and insomnia, which may result in daytime somnolence as well. Perhaps more frequently, the clinician is called upon to treat the medical consequences of the disorder as it worsens during sleep.

Patients with *fibrositis syndrome* complain of chronic fatigue and diffuse musculoskeletal pain, and physical examination is said to reveal specific points of tenderness over certain muscles. Sleep is characteristically described as unrefreshing, and a polysomnographic finding termed "alpha-delta sleep" has been reported.

Sleeping sickness is a chronic meningoencephalomyelitis endemic to tropical Africa and caused by the protozoan *Trypanosoma brucei.* Patients develop sleep fragmentation and hypersomnolence along with profound abnormalities of the sleep electroencephalogram; when untreated, the disease is lethal.

Elements of the Sleep Disorders Consultation

Given the broad scope of sleep disorders medicine, and the frequent interaction between sleep/wake disorders and medical or psychiatric illness, it follows that part of the sleep disorders consultation should include a standard, thorough medical history and physical examination. In addition, several aspects of the consultation specific to sleep medicine deserve mention.

History

Sleep Habits and Schedule. A thorough understanding of the patient's sleep-wake schedule and bedtime habits can be surprisingly helpful. Timing and variability of bedtime and arising should be recorded for both regular workdays and days off, and the patient should be asked to estimate the average amount of sleep obtained (including naps), accounting for time to sleep onset and any customary awakenings during sleep. It is surprising how many patients, presenting with a complaint of excessive sleepiness, are found to have insufficient sleep syndrome as at least a component of their problem. An almost invariable response to the time pressures of modern life is to choose to make less time available for sleep, and it is astonishing how ignorant the general public (let alone the medical community) can be concerning the effects of sleep deprivation (15). Also, the estimate of time to sleep onset can be helpful in other ways: Patients who complain of sleep onset insomnia but fall asleep in 30 minutes or less should be reassured that their sleep latency is, in fact, normal; patients who regularly fall asleep "as their head hits the pillow" may actually be excessively sleepy.

Any variability of the sleep schedule should be ascertained, especially in the patient complaining of insomnia. A regular sleep/wake schedule is exceedingly important in maintaining the accuracy of the internal circadian rhythm of sleepiness, and the patient who habitually "sleeps in" on weekends or after sleepless nights, or naps excessively, can forfeit the regular nightly onset of sleepiness that helps one fall asleep. In addition, a careful delineation of the sleep/wake schedule is instrumental in diagnosing many of the circadian rhythm disorders, particularly shift work sleep disorder, irregular or non-24-hour sleep-wake disorder, and the sleep phase syndromes. In these situations, a 14-day sleep log detailing time in bed and estimates of when the patient is asleep can be of great utility.

Bedtime habits are frequently illuminating, especially when the patient complains of insomnia. Use of alcoholic or caffeine-containing beverages, sedative-hypnotics, or illicit drugs should be detailed. The consultant should inquire as to violations of the rules of sleep hygiene detailed in Table 4, such as exercising too close to bedtime, noise in the bedroom, or activities in bed not conducive to sleep. A common error practiced by the patient with psychophysiologic insomnia is to keep a clock within easy view; many such patients regularly check the time as the sleepless night progresses, and become increasingly anxious when the clock reminds them that they have not been sleeping.

Phenomena Associated with Sleep. The patient can sometimes supply information concerning events during the night that point to certain sleep disorders. More frequently, the bed partner or other observer will be in a position to describe details of sleep phenomena that are of great diagnostic utility; if at all possible, the bed partner or observer should be questioned before a final opinion is rendered. Recurrent awakenings from sleep can be indicative of sleep disorders that cause sleep fragmentation, such as sleep apnea syndromes or periodic limb movement disorder. Habitual early morning awakening is frequent in depressive illness. In the case of sleep apnea, the patient may report episodes of air hunger or choking upon awakening, especially when central sleep apnea exists (16). Headache or nausea, dry mouth or sore throat, palpitations, or acid brash upon awakening may also suggest sleep apnea, as can nocturia and enuresis. Often the bed partner may remark on resuscitative snoring, an intermittent nature to the snoring, or vividly describe apneas, thus providing strong evidence for sleep apnea. In patients with periodic limb movement disorder, the bed partner may actually observe the characteristic limb movements, or the patient may report the symptoms of associated restless legs syndrome at bedtime or during nighttime awakenings.

The reports of bed partners or other observers are especially important when parasomnias are being investigated. They can provide graphic descriptions of confusional arousals, sleepwalking, sleep terrors, REM behavior disorder episodes, seizures, or dystonic movements that strongly implicate a specific diagnosis. The timing of these episodes (early in the night vs last half of the night) is extremely helpful in differentiating an arousal disorder from a parasomnia associated with REM sleep.

Excessive Daytime Sleepiness. This may be a chief complaint, or may only be elicited as a problem as the consultation progresses. Sleepiness, like pain, is a subjective complaint, and many patients may greatly underestimate its severity or will not appreciate whether sleepiness is present at all; it can also be masked by activity or suppressed by chronic use of stimulants such as caffeine. At times, only another observer (most frequently a spouse) will provide the evidence of unintended dozing that indicates hypersomnolence. In addition, many patients confuse the symptoms of fatigue and sleepiness; fatigue is perhaps ubiquitous in today's society, whereas sleepiness is a very specific symptom suggesting a sleep/wake disorder. Arguably the best way to screen for the presence of sleepiness by history is to ask whether the patient habitually falls asleep during certain situations. The situations may range from those that are frequently sleep inducing (darkened lecture halls, reading in the evening) to those that would be associated with dozing only in the sleepiest of individuals (driving, attending a sporting event other than baseball). This method of quantiating sleepiness by history has been codified by Johns in the form of the Epworth Sleepiness Scale (**Table 5**), which consists of asking the likelihood of dozing (never, slight, moderate, or high) during a list of eight specific situations. A score is computed by assigning a number to each likelihood (0 to 3) and summing the numbers recorded for each situation. Johns

TABLE 5

The Epworth Sleepiness Scale

How likely are you to doze off or fall asleep in the following situations, in contrast to feeling just tired? This refers to your usual way of life in recent times. Even if you have not done some of these things recently try to work out how they would have affected you. Use the following scale to choose the *most appropriate number* for each situation:

0 = would *never* doze
1 = *slight* chance of dozing
2 = *moderate* chance of dozing
3 = *high* chance of dozing

Situation	Chance of Dozing
Sitting and reading	
Watching TV	
Sitting, inactive in a public place (e.g., a theatre or a meeting)	
As a passenger in a car for an hour without a break	
Lying down to rest in the afternoon when circumstances permit	
Sitting and talking to someone	
Sitting quietly after a lunch without alcohol	
In a car, while stopped for a few minutes in the traffic	

From: Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991; 14:540–545, with permission.

found that scores of 11 or greater were consistently present in patients with disorders of excessive daytime sleepiness and that the score was significantly correlated with the objective measurement of sleepiness by multiple sleep latency testing (17).

REM-Intrusive Phenomena. REM-intrusive phenomena (hypnagogic hallucinations, sleep paralysis, cataplexy) can be important clues to the diagnosis of narcolepsy syndrome. However, they are not invariably present: up to 30% of patients with documented narcolepsy never develop cataplexy, and as many as 75% never complain of hypnagogic hallucinations or sleep paralysis (18). In addition, other disorders associated with sleep fragmentation and REM deprivation can lead to increased "REM pressure" and REM-intrusive phenomena; in particular, hypnagogic hallucinations have been reported in obstructive sleep apnea syndrome (19). Despite these caveats, evidence of REM-intrusive phenomena should be sought in patients with complaints of excessive daytime sleepiness. A history of hypnagogic hallucinations (usually visual or auditory, but occasionally tactile or olfactory (20)) or sleep paralysis is elicited fairly readily, as these symptoms are striking in nature. Cataplectic episodes, however, may be considerably more subtle. One should inquire, not only about full-blown episodes resulting in complete atonia and collapse, but also about brief weakness in the knees or neck, or a momentary sagging of the jaw or mumbling of speech. An increase in dream frequency, dreaming during brief naps, or dreaming early in the night may also be suggestive of increased "REM pressure."

Past Medical History. A comprehensive past medical history may reveal medical or psychiatric illness that can have profound effects on sleep (e.g., those listed in Table 3). In patients suspected of having sleep apnea, a host of neurologic disorders (21) as well as conditions affecting the upper airway (22) (e.g., allergic rhinitis) may contribute to the pathogenesis, and there is a strong association between obstructive sleep apnea and hypertension (23).

A complete list of medications used (both prescribed and over-the-counter) is essential, as many drugs can have profound effects on sleep and wakefulness. For the same reason, the consultant should also inquire about illicit drug use and should have the patient estimate his or her intake of alcoholic beverages, caffeine-containing beverages, and tobacco.

Family History. A number of sleep disorders exhibit a familial tendency, and a family history of certain symptoms or specific disorders can be a valuable clue for the consultant. A particularly strong familial association exists for narcolepsy and hypersomnia (20, 24) and the arousal disorders (25); genetic predisposition may also be a factor in insomnia (26) and obstructive sleep apnea (27).

Physical Examination

The physical examination in sleep disorders medicine may properly be said to begin with a trip to the waiting room to retrieve the patient: a not-inconsiderable number of patients with hypersomnolence will be found asleep while waiting to be called, and some may even exhibit obstructive respiratory events, thus virtually announcing their diagnosis. In the case of an in-patient, quiet entrance to the hospital room may reward the consultant with the opportunity to observe the sleeping patient and make a similar observation. At the very least, a patient asleep in the waiting room corroborates any complaint of excessive sleepiness. Other aspects of general observation can be important in the awake patient: a sad affect or tearfulness will raise the suspicion that depressive illness may be involved in a sleep disturbance, and obesity, craniofacial abnormality, or plethora might suggest the possibility of a sleep apnea syndrome.

Evaluation of the upper airway is particularly important in patients suspected of having obstructive sleep apnea syndrome. Although many patients with sleep apnea will have no obvious upper airway abnormality, the presence of nasal obstruction or signs of allergic rhinitis (28), narrowing of the oropharyngeal lumen or palatal edema and erythema (29), tonsillomegaly, macroglossia, or obvious upper airway rumors or lesions will raise the possibility of that diagnosis. Some studies have implied that neck circumference is an important predictor of obstructive sleep apnea, but no simple guidelines are available (30).

The pulmonary examination may uncover evidence of bronchial asthma or chronic obstructive pulmonary disease suggesting the presence of sleep fragmentation due to these respiratory disorders. Signs of pulmonary hypertension or *cor pulmonale* (pedal or presacral edema, neck vein distention, right-sided gallop, accentuated P_2, murmurs of pulmonary or tricuspid regurgitation) or cardiac examination may suggest the sleep apnea syndromes, central alveolar hypoventilation, or chronic lung disease leading to disordered sleep. The abdominal examination can reveal hepatomegaly, raising the suspicion of alcohol abuse. The neurologic examination, including mental status testing, can yield important clues to the presence of a dementing illness or other neurologic disorder leading to sleep disturbance.

Laboratory Testing

Standard Laboratory Testing. Information with respect to several sleep disorders can be obtained from certain standard laboratory tests. The complete blood count may reveal polycythemia, which suggests the possibility of a sleep apnea syndrome or central alveolar hypoventilation syndrome. Similarly, elevated total CO_2 as determined from a serum chemistry panel may be an indicator of chronic respiratory acidosis from central alveolar hypoventilation. Of course, confirmation should be sought by a determination of arterial blood gases. The electrocardiogram and chest radiograph can provide evidence for pulmonary hypertension or *cor pulmonale,* which may also be sequelae of sleep-disordered breathing. Although there is an association between hypothyroidism and obstructive sleep apnea syndrome, the diagnostic yield from performing thyroid-stimulating hormone assays on all patients suspected of sleep apnea appears to be low (31). Pulmonary function testing, particularly spirometry, can be useful in screening for the presence of other pulmonary disorders that can lead to hypercapnia or sleep disordered breathing (e.g., chronic obstructive lung disease or bronchial asthma). However, despite an initial burst of enthusiasm, spirometric evidence of variable upper airway obstruction (flattening of the inspiratory limb of the flow-volume loop or "saw-toothing" of either the inspiratory or expiratory limbs) has proven to be of little use as a diagnostic tool (32).

Sleep Laboratory Testing. Two types of sleep testing services are generally available in the sleep laboratory: *overnight polysomnography* and the *multiple sleep latency test.* Overnight polysomnography is defined as the collection of multiple physiologic signals during sleep; such a study, performed with a polysomnographic technologist in attendance, is currently the reference standard for the diagnosis of many sleep disorders. The signals collected include those necessary to stage sleep (electroencephalogram, electrooculogram, submental electromyogram), identify apneas (respiratory effort, airflow, oxyhemoglobin saturation), detect limb movements (electromyograms, usually of the anterior tibialis muscles), and monitor cardiac rhythm (electrocardiogram). Additional signals may be included when specific sleep disorders are suspected (e.g., esophageal pH for sleep-related gastroesophageal reflux). Video recording is frequently included, especially when a parasomnia is suspected. The testing is performed during at least 6 to 8 hours of nocturnal sleep; a patient with a different sleep schedule (e.g., night-shift worker) will be tested during the time of their usual major sleep period.

A variety of time- and labor-saving alternatives have been proposed over the years to replace attended overnight polysomnography. Daytime ("nap") studies are not encouraged: such studies provide no information relating to the patient's usual sleep architecture, and may underestimate the severity of sleep apnea because REM sleep (when apneas are typically most severe) is rarely observed (33). At present, unattended home recordings are felt to have limited utility. These studies suffer from the lack of a trained technologist who can correct the technical problems that invariably occur during a prolonged recording, institute treatment when apneas are complicated by malignant arrhythmias, directly observe and interpret nighttime behaviors, and educate the patient. Current standards of practice limit unattended polysomnography to situations where severe obstructive sleep apnea syndrome is likely and laboratory polysomnography is not immediately available; when patients cannot be studied in the laboratory; or for the evaluation of therapeutic response after diagnosis and proper therapy have been established by laboratory polysomnography (34).

In 1982, the majority of patients undergoing polysomnography at sleep disorders centers were found to have a disorder associated with excessive daytime sleepiness (9). This symptom, in the absence of compelling evidence for insufficient sleep syndrome, probably remains the most common indication for polysomnographic testing. In addition, standards of practice have been published summarizing the indications for overnight polysomnography with respect to cardiopulmonary disorders (35) and insomnia (36). These indications are summarized in **Tables 6 and 7,** respectively.

The multiple sleep latency test (MSLT) is a daytime procedure used to objectively measure sleepiness and to detect any tendency for premature onset of REM sleep (37, 38). The study must always be preceded by overnight polysomnography, so that the MSLT can be interpreted in light of the quality of sleep received. At a minimum, an electroencephalogram, electrooculograms, and submental electromyogram are collected to determine the presence of wakefulness or sleep, and to stage sleep. Starting within 90 to 180 minutes of awakening from the polysomnogram, the patient is presented with a series of five opportunities to nap, each separated by 2 hours of enforced wakefulness. The study is performed in a darkened

TABLE 6

Indications for Polysomnography in Cardiopulmonary Disorders

Excessive daytime sleepiness or sleep maintenance insomnia

Snoring or obesity only when accompanied by other symptoms or risk factors for sleep apnea

Nocturnal cyclic bradytachyarrhythmias, nocturnal abnormalities of atrioventricular conduction, or ventricular ectopy during sleep that appears increased relative to wakefulness

Disturbances of respiratory control when awake pCO_2 is greater than 45 Torr, or when complicated by pulmonary hypertension, polycythemia, disturbed sleep, morning headaches, or daytime somnolence and fatigue

Chronic obstructive pulmonary disease when complicated by pulmonary hypertension, right heart failure, or polycythemia despite awake pO_2 of greater than 55 Torr

Restrictive ventilatory disorders complicated by chronic hypoventilation, polycythemia, pulmonary hypertension, disturbed sleep, morning headaches, or daytime somnolence and fatigue

Adapted from: American Thoracic Society Consensus Conference, Phillipson EA and Remmers JE, Chairmen. Indications and standards for cardiopulmonary sleep studies. Am Rev Respir Dis 1989; 139:559–568, with permission.

TABLE 7

Indications for Polysomnography in Insomnia

Polysomnography is indicated in the evaluation of suspected sleep-related breathing disorders and periodic limb movement disorder, conditions that are most likely to cause insomnia in middle-aged or elderly patients

Polysomnography is indicated when the initial diagnosis is uncertain or when the treatment, either behavioral or pharmacologic, is unsuccessful

Polysomnography is indicated for the diagnosis of precipitous arousals or violent behavior in sleep when the clinical diagnosis is uncertain

Polysomnography may be indicated for the diagnosis of persistent circadian rhythm disorders, such as delayed-sleep-phase syndrome, when the clinical diagnosis is unclear

Polysomnography is not indicated for the routine evaluation of transient or chronic insomnia

Adapted from: American Sleep Disorders Association Standards of Practice Committee. Practice parameters for the use of polysomnography in the evaluation of insomnia. Sleep 1995; 18:55–57, with permission.

room and in an environment conducive to sleep, the object being to measure the time from "lights out" to the appearance of any stage of sleep. The patient is afforded 20 minutes to fall asleep; an additional 15 minutes is allowed, if sleep occurs within this time, to determine whether REM sleep will appear.

Analysis of the MSLT results in two indices. The *mean sleep latency,* defined as the average time required for the patient to fall asleep, is used as an indicator of the overall sleepiness of the individual. A value over 10 minutes is considered normal; under 5 minutes indicates pathologic sleepiness; and a value between 5 and 10 is considered marginal or indeterminate by some, or as indicating a milder degree of sleepiness by others. The number of *sleep onset REM periods* (simply the number of naps during which REM sleep occurred within 15 minutes of sleep onset) is used in the diagnosis of narcolepsy: a mean sleep latency of 5 minutes or less in combination with two sleep onset REM periods, in the absence of evidence of significant REM deprivation on the preceding night's polysomnogram, is considered diagnostic of narcolepsy. The MSLT is most often used in the diagnosis of narcolepsy or other disorders of excessive daytime sleepiness; to document the effect of treatment (e.g., in obstructive sleep apnea syndrome) on objective sleepiness; or to substantiate

TABLE 8

Indications for Multiple Sleep Latency Testing

Narcolepsy: An MSLT is indicated for all patients suspected of narcolepsy to confirm the diagnosis and to determine the severity of sleepiness and should be performed before commencing treatment with stimulant medications

Obstructive sleep apnea syndrome: (a) The MSLT is indicated in patients with mild to moderate obstructive sleep apnea syndrome who complain of moderate to severe sleepiness (b) The MSLT may be indicated in patients with moderate to severe obstructive sleep apnea syndrome, especially if severe sleepiness in unappreciated or denied

Other causes of excessive sleepiness: An MSLT is indicated in the evaluation of patients suspected of having idiopathic hypersomnia, periodic limb movement disorder or when the cause of excessive sleepiness is unknown

Insomnia: The MSLT is indicated in the evaluation of the complaint of insomnia when the presence of moderate to severe excessive sleepiness is suspected

Circadian rhythm disorders: The MSLT may be useful in documenting sleepiness in some circadian rhythm sleep disorders, but adequate scientific validation is not yet available

Assessment of treatment effects: The MSLT is indicated to assess the response to treatment following effective therapy for disorders that cause sleepiness when an additional sleep disorder that produces sleepiness is suspected, or if confirmation of relief of sleepiness is required to ensure occupational safety

Repeat MSLT testing: Repeat MSLT testing is indicated in the following situations: (a) when the initial test is believed to be an invalid representation of the patient's status; (b) when ambiguous or uninterpretable MSLT findings occur; (c) when the response to treatment needs to be ascertained; and (d) when more than sleep disorder is suspected

Adapted from: American Sleep Disorders Association Standards of Practice Committee, Thorpy MJ, Chairman. The clinical use of the multiple sleep latency test. Sleep 1992; 15:268–276, with permission.

complaints of sleepiness in a broad variety of sleep disorders (Table 8).

A variant of the MSLT, the *maintenance of wakefulness test* (*MWT*), is also performed in some institutions. The MWT is performed in exactly the same manner as the MSLT, except the patient is asked to try to stay awake, rather than fall asleep, during each of the 20-minute testing periods. Mean sleep latencies derived from the MWT do not closely correlate with those from the MSLT, and it has been suggested that the MWT may be a better test of the ability to stay alert for tasks such as driving (39). The use of the MWT is somewhat hampered by the relative lack of accepted normative values compared to the wealth of such data available for the MSLT.

Management

Space does not permit a comprehensive exposition on treatment of the broad range of sleep disorders. However, a number of specific issues concerning common sleep disorders deserve mention.

Insomnia. Insomnia is such a frequent complaint (3) that the primary care physician should be expected to develop considerable skills in its management. The record, as reflected in research studies on this subject, appears to be somewhat uneven. A national survey of physicians' prescribing patterns in insomnia suggests that indiscriminate use of hypnotic medication is not a significant problem (40). However, there are data suggesting that primary care for insomnia is likely to be incomplete and cursory. For instance, a recent study by Everitt and colleagues (41) tested the diagnostic approach to insomnia used by physicians in general practice by presenting them with standard case vignettes (41). These practitioners asked an average of only 2.5 questions in eliciting further history, slightly less than half of them obtained a sleep history, and about the same number prescribed medication for the complaint rather than pursuing the many nonpharmacologic treatment modalities available.

Acute, self-limited insomnia related to a catastrophic life event can often best be treated by the short-term use of hypnotic medication, especially when daytime function is significantly impacted (42). Pharmacologic therapy also plays an important role in the more than one third of patients who are found to have a psychiatric illness underlying their insomnia (9, 43). However, the majority of these patients have a mood disorder; monotherapy with antidepressant medication is generally more appropriate than hypnotic use, and referral to a psychiatric professional is certainly encouraged.

A substantial number of the remaining patients have psychophysiologic insomnia, sleep state misperception, or idiopathic insomnia (9, 43). The care of these patients is often time-intensive, and the sleep medicine consultant, along with the resources of the sleep disorders center, can be of great assistance to the primary care practitioner. In general, management revolves around the precepts of sleep hygiene outlined in Table 4, which can take considerable, ongoing interaction between the physician and patient to accomplish. Training in relaxation techniques, including self-hypnosis and biofeedback (44), are usually indicated and commonly can be arranged through the sleep disorders center. Stimulus control therapy attempts to break the association between the sleeping environment and wakefulness by removing the patient from the bedroom when not actually asleep or drowsy (45). Finally, hypnotic medication (primarily benzodiazepines or, more recently, zolpidem) should be reserved for adjunctive use in these patients (46). Current practice is to use the lowest dose possible, on an intermittent basis only (e.g., every third night), and with due regard for the pharmacokinetic characteristics of the agent chosen (47).

Obstructive Sleep Apnea Syndrome. Management of the sleep apnea syndromes has recently been reviewed in this journal (48). Treatment is usually instituted for moderate or severe disease whether or not symptoms are present, and in mild cases when daytime sleepiness or other symptoms are felt to be related to the disorder. The mainstay of treatment continues to be nasal continuous positive airway pressure (C-PAP); however, compliance with this therapy is a concern (49). Mechanical problems with the mask, local irritation from the airflow or mask apparatus, and nasal symptoms (rhinorrhea, sneezing) often complicate the patient's attempts to use this modality. Ongoing communication between the patient and sleep medicine consultant is often helpful in resolving these problems to enhance compliance. Other modalities of treatment are available when the patient will not use nasal C-PAP: conventional uvulopalatopharyngoplasty, oral appliances, and weight loss all have their role in therapy. However, treatment outcomes are generally less predictable compared with nasal C-PAP, and patients must be closely followed to verify ongoing treatment efficacy; this will often include repeat overnight polysomnography. Another related issue that is often (or should be) the purview of the sleep medicine consultant concerns the patient requesting laser-assisted uvulopalatoplasty (LAUP) for socially disruptive snoring. A significant number of these patients will prove to have obstructive sleep apnea syndrome; because sleep apnea, unlike snoring, is not adequately treated by LAUP (50), and cannot be predicted reliably from the preoperative history and physical examination, sleep disorders center evaluation for this disorder is mandatory before LAUP is performed (51).

Narcolepsy. Treatment for narcolepsy generally involves the use of central nervous system stimulants (pemoline, methylphenidate, methamphetamine hydrochloride) to relieve daytime sleepiness, and REM suppressant medication (tricyclic antidepressants, selective serotonin reuptake inhibitors, and possibly monoamine oxidase inhibitors) to ameliorate cataplexy, sleep paralysis, and hypnagogic hallucinations. An adjunct to pharmacologic therapy involves frequent daytime naps: a brief, 15-minute nap can significantly increase sleep latency in such patients, although the duration of this effect is relatively brief (52). The physician can often be of assistance by interceding for the patient with an employer who will not allow napping. Participation in a support group is often beneficial, as the narcoleptic is frequently perceived as being “lazy” by family, friends, and employers; interacting with other narcoleptics who have worked through similar experiences can do much for the patient’s morale. Finally, as many as 30% of patients with narcolepsy may have disturbed nocturnal sleep attributable to sleep apnea or periodic limb movement disorder, and suitable treatment of these complicating conditions can improve daytime function (53).

Concluding Remarks

The enormous prevalence of sleep disorders ensures that sleep/wake complaints will increasingly be brought to the attention of the primary care practitioner. The informed generalist can be quite effective in diagnosing and treating a number of sleep disorders, and should also know when to take advantage of the consultative, testing, and educational services of the local sleep disorders centers. Finally, it is incumbent upon both the referring physician and the consultant to communicate liberally with each other, both verbally and in writing. A surprising number of consultants answer questions that were not asked simply because the referring physician did not convey the reason for the consultation; conversely, the consultant must adequately inform and educate the referring physician if the patient is to derive full benefit from the encounter (54). It is to be hoped that the time pressures and paradigm shifts being imposed on physicians by the rapid changes in the medical marketplace will not impair this essential interchange.

References

1. Luke RD, Thomson MA. Utilization of within-hospital services: a study of the effects of two forms of group practice. Med Care 1980; 18:219–227.
2. Dement WC, Mitler MM. It’s time to wake up to the importance of sleep disorders. JAMA 1993; 269:1548–1550.
3. Bixler EO, Kales A, Soldatos CR et al. Prevalence of sleep disorders in the Los Angeles metropolitan area. Am J Psychiatry 1979; 136:1257–1262.
4. Mitler MM, Carskadon MA, Czeisler CA et al. Catastrophes, sleep, and public policy: consensus report. Sleep 1988; 11:100–109.
5. Dement WC. Wake up, America! Education about sleep deprivation and sleep disorders. Presented at the National Conference on Education About Sleep, Columbia Presbyterian School of Medicine, November 20–21, 1992, New York.
6. Rosen RC, Rosekind M, Rosevear C et al. Physician education in sleep and sleep disorders: a national survey of U.S. medical schools. Sleep 1993; 16:249–254.
7. Schaffner F. Consultation in hepatology. Mt Sinai J Med 1994; 61:509–511.
8. Diagnostic Classification Steering Committee, Thorpy MJ, Chairman. International classification of sleep disorders: diagnostic and coding manual. Rochester, MN: American Sleep Disorders Association, 1990.
9. Coleman RM, Roffwarg HP, Kennedy SJ et al. Sleep-wake disorders based on a polysomnographic diagnosis. JAMA 1982; 247:997–1003.
10. Young T, Palta M, Dempsey J et al. The occurrence of sleep-disordered breathing among middle-aged adults. N Engl J Med 1993; 328:1230–1235.
11. He J, Kryger MH, Zorick FJ et al. Mortality and apnea index in obstructive sleep apnea. Experience in 385 male patients. Chest 1988; 94:9–14.
12. Palomaki H, Partinen M, Erkinjuntti T, Kaste M. Snoring, sleep apnea syndrome, and stroke. Neurology 1992; 42(suppl):75–82.
13. Stoohs RA, Guilleminault C, Itoi A, Dement WC. Traffic accidents in commercial long-haul truck drivers: the influence of sleep-disordered breathing and obesity. Sleep 1994; 17:619–623.
14. Devinsky O, Ehrenberg B, Barthlen GM et al. Epilepsy and sleep apnea syndrome. Neurology 1994; 44:2060–2064.
15. Dement WC, Gelb M. Somnolence: its importance in society. Neurophysiol Clin 1993; 23:5–14.
16. Bradley TD, McNicholas WT, Rutherford R et al. Clinical and physiologic heterogeneity of the central sleep apnea syndrome. Am Rev Respir Dis 1986; 134:217–221.
17. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991; 14:540–545.
18. Manfredi RL, Brennan RW, Cadieux RJ. Disorders of excessive sleepiness: narcolepsy and hypersomnia. Semin Neurol 1987; 7:250–258.
19. Guilleminault C, van den Hoed J, Mitler MM. Clinical overview of the sleep apnea syndromes. In: Guillemi-

nault C, Dement WC, eds. Sleep apnea syndromes. New York: Alan R. Liss, 1978:1–12.

20. Kales A, Cadieux RJ, Soldatos CR et al. Narcolepsy-cataplexy. I. Clinical and electrophysiologic characteristics. Arch Neurol 1982; 39:164–168.
21. Brown LK. Sleep-disordered breathing in neurologic disease. Clin Pulmon Med In press.
22. Brown LK. Sleep apnea syndromes: overview and diagnostic approach. Mt Sinai J Med 1994; 61:99–112.
23. Hla KM, Young TB, Bidwell T et al. Sleep apnea and hypertension. A population-based study. Ann Intern Med 1994; 120:382–388.
24. Honda T, Asaka A, Tanimura M, Furosho T. A genetic study of narcolepsy and excessive daytime sleepiness in 308 families with a narcolepsy or hypersomnia proband. In: Guilleminault C, Lugaresi E, eds. Sleep/wake disorders: natural history, epidemiology, and long-term evolution. New York: Raven Press, 1983.
25. Kales A, Soldatos CR, Bixler EO et al. Hereditary factors in sleepwalking and night terrors. Br J Psychiatry 1980; 137:111–118.
26. Heath AC, Kendler KS, Eaves LJ, Martin NG. Evidence for genetic influences on sleep disturbance and sleep pattern in twins. Sleep 1990; 13:318–335.
27. Strohl KP, Saunders NA, Feldman NT, Hallet M. Obstructive sleep apnea in family members. N Engl J Med 1978; 299:969–972.
28. McNicholas WT, Tarlo S, Cole P et al. Obstructive apneas during sleep in patients with seasonal allergic rhinitis. Am Rev Respir Dis 1982; 126:625–628.
29. Shepard JW Jr, Gefter WB, Guilleminault C et al. Evaluation of the upper airway in patients with obstructive sleep apnea. Sleep 1991; 14:361–371.
30. Flemons JW, Whitelaw WA, Brant R, Remmers JE. Likelihood ratios for a sleep apnea clinical prediction rule. Am J Respir Crit Care Med 1994; 150:1279–1285.
31. Lin C-C, Tsan K-W, Chen P-J. The relationship between sleep apnea syndrome and hypothyroidism. Chest 1992; 102:1663–1667.
32. Krieger J, Weitzenblum E, Vandevenne A et al. Flow-volume curve abnormalities and obstructive sleep apnea syndrome. Chest 1985; 87:163–167.
33. Sériès F, Cormier Y, La Forge J. Validity of diurnal sleep recording in the diagnosis of sleep apnea syndrome. Am Rev Respir Dis 1991; 143:947–949.
34. American Sleep Disorders Association Standards of Practice Committee, Thorpy M, Chairman. Practice parameters for the use of portable recording in the assessment of obstructive sleep apnea. Sleep 1994; 17:372–377.
35. American Thoracic Society Consensus Conference, Phillipson EA and Remmers JE, Chairmen. Indications and standards for cardiopulmonary sleep studies. Am Rev Respir Dis 1989; 139:559–568.
36. American Sleep Disorders Association Standards of Practice Committee, Thorpy M, Chairman. Practice parameters for the use of polysomnography in the evaluation of insomnia. Sleep 1995; 18:55–57.
37. Association of Sleep Disorders Centers Task Force on Daytime Sleepiness, Carskadon MA, Chairman. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. Sleep 1986; 9:519–524.
38. American Sleep Disorders Association Standards of Practice Committee, Thorpy MJ, Chairman. The clinical use of the multiple sleep latency test. Sleep 1992; 15:268–276.
39. Sangal RB, Thomas L, Mitler MM. Maintenance of wakefulness test and multiple sleep latency test. Measurement of different abilities in patients with sleep disorders. Chest 1992; 101:898–902.
40. Mellinger GD, Balter MB, Uhlenhuth EH. Insomnia and its treatment. Arch Gen Psychiatry 1985; 42:225–232.
41. Everitt DE, Avorn J, Baker MW. Clinical decision-making in the evaluation and treatment of insomnia. Am J Med 1990; 89:357–362.
42. Dement WC. The proper use of sleeping pills in the primary care setting. J Clin Psychiatry 1992; 53(suppl): 50–56.
43. Buysse DJ, Reynolds CF III, Kupfer DJ et al. Clinical diagnoses in 216 insomnia patients using the international classification of sleep disorders (ICSD), DSM-IV and ICD-10 categories: a report from the APA/NIMH DSM-IV field trial. Sleep 1994; 17:630–637.
44. Hauri P. Treating psychophysiological insomnia with biofeedback. Arch Gen Psychiatry 1981; 752–758.
45. Bootzin RR, Nicassio PM. Behavioral treatments for insomnia. In: Hersen M, Eissler R, Miller P, eds. Progress in behavior modification. Vol. 6. New York: Academic Press, 1978:1–45.
46. National Institutes of Health. Consensus development on drugs and insomnia conference. JAMA 1984; 251:2410–2414.
47. Bliwise DL. Treating insomnia: pharmacological and nonpharmacological approaches. J Psychoactive Drugs 1991; 23:335–341.
48. Rapoport DM. Treatment of sleep apnea syndromes. Mt Sinai J Med 1994; 61:123–130.
49. Engleman HM, Martin SE, Douglas NJ. Compliance with CPAP therapy in patients with the sleep apnoea/hypopnoea syndrome. Thorax 1994; 49:263–266.
50. Krespi YP, Keidar A, Khosh MM et al. The efficacy of laser-assisted uvulopalatoplasty (LAUP) in the management of obstructive sleep apnea syndrome (OSA) and upper airway resistance syndrome (UARS). Head Neck Surg 1994; 5:235–243.
51. American Sleep Disorders Association Standards of Practice Committee, Thorpy M, Chairman. Practice parameters for the use of laser-assisted uvulopalatoplasty. Sleep 1994; 17:744–748.
52. Roehrs T, Zorick F, Wittig R et al. Alerting effects of naps in patients with narcolepsy. Sleep 1986; 9:194–199.
53. Wittig R, Zorick F, Piccione P et al. Narcolepsy and disturbed nocturnal sleep. Clin Electroencephalogr 1983; 14:130–134.
54. Lee T, Pappius EM, Goldman L. Impact of inter-physician communication on the effectiveness of medical consultations. Am J Med 1983; 74:106–112.

Evolution of Cardiologic Consultation

Leslie A. Kuhn, M.D.

In my college days Webster's dictionary defined a consultant as "one who gives professional advice or services" in the field of his or her special knowledge. Despite many advances and changes in cardiology since then, that dictionary definition has not changed. What has changed is the scope of knowledge required and the number of consultants necessary to achieve optimal management in many clinical situations.

The complexity of cardiologic consultations has increased proportionally with advances in knowledge in the field. When I trained in cardiology only congenital heart disease and some valvular diseases could be treated surgically, pharmacologic agents were few, and the management of acute myocardial infarction was limited to 4 to 6 weeks of bed rest accompanied by words of encouragement. There was no electrophysiology, echocardiography, lipid analysis, sophisticated appreciation of "risk factors," coronary arteriography, or nuclear cardiology, all of which have changed the scope of consultation and the knowledge required of the consultant.

At that time, consultants were generally senior physicians brought in to reassure and console the patient and members of the family because relatively little could be done to improve management of the patient.

In contrast, today several consultants may be required to provide optimal advice, including electrophysiologists, who are experts in the diagnosis and treatment of complex arrhythmias; lipid experts; echocardiographers; nuclear cardiologic specialists; cardiac catheterizers; and, of course, those who provide therapy directly, interventional cardiologists and cardiac surgeons. Many of these are younger physicians who have achieved considerable special knowledge in their area of expertise. Their opinions are usually of great value but the management of the patient usually requires that one physician collate all the information and apply it to the patient's particular needs and psychological structure, taking other medical and social factors into account. The coordinator is a clinical cardiologist working with the patient's primary physician, usually an internist.

Patients now possess much more knowledge (and misinformation) about disease than they once did because of repeated exposure to information in newspapers, television, and magazines. "Health" is probably a primary interest of many people. At one time a consultant's opinion generally would be accepted with relatively little discussion, but in today's environment that is not possible or desirable.

The downside of this public information about disease is its fragmentary and naive nature; it is not presented in a context of appreciation for statistical concepts and distinctions are not made between a testimonial or anecdote and valid data. I am amazed how often intelligent people have no appreciation of what data are believable and what are not. Medicine may be one of the few fields of specialized knowledge in which people without any knowledge or scientific training give opinions on appropriate therapy. What is more amazing is that these opinions are accepted so frequently. In selecting a cardiac surgeon, for example, a patient may listen to knowledgeable specialists in the field and still select an individual or institution recommended by a relative, friend, or employer or base the selection on an advertisement for a particular institution.

The essence of consultation still has not changed. As in the practice of all medicine it is a mixture of art and science. Patients seek help in managing a specific problem or many problems,

---

From the Division of Cardiology, Department of Medicine, The Mount Sinai School of Medicine, New York, NY 10029. Address reprint requests to Leslie A. Kuhn, M.D., Division of Cardiology, Department of Medicine, The Mount Sinai School of Medicine, 1 Gustave L. Levy Place, New York, NY 10029-6574.

and the kind of help they seek may vary enormously.

There may be a problem with no definite concrete answer—for example, when or if to operate on an asymptomatic person with advanced valvular disease. This is a common clinical problem, the subject of many published discussions, yet there is no definitive, absolutely correct answer at this time. Existing data indicate that aortic or mitral valve replacement is less successful if left ventricular functional deterioration has occurred, and various echocardiographic or nuclear cardiologic criteria are used to establish such dysfunction. Of course, surgeons are interested in operating on the patient when the risk is lowest and the outcome likely to be best, and many asymptomatic patients fall into this category. Nevertheless, I have been reluctant to advise surgery for an asymptomatic patient because every procedure entails some mortality and morbidity. In such a situation, knowledge of the natural history of the condition without surgery is helpful and such knowledge is augmented greatly by experience. In aortic regurgitation, for example, the patient may have a long history of an asymptomatic stable state and a productive life with no substantial limitation of activities, including competitive sports, even though echocardiographic or nuclear cardiologic criteria of left ventricular dysfunction (necessitating valve replacement according to the current literature) have been met.

In analyzing recommendations for early surgical intervention in the asymptomatic patient, the historical and the projected duration of the asymptomatic state is usually not considered nor is the possibility of beneficial technical advances in surgery that may occur during the waiting period. Many patients will not ever require surgery.

All these considerations must be discussed with the patient in helping him or her to make the decision. If the patient or the consultant is not "ready" for surgery, further clinical observation with appropriate periodic testing is usually the wisest course and can generally be followed safely. The patient may have strong feelings about whether he or she wants an intervention performed and may seek many opinions until finding one that supports his or her feelings.

An issue may be clear to most physicians, but the patient or family may want a confirmatory opinion. An example of this is an active individual with limiting angina pectoris, or perhaps a mildly symptomatic person who has had a stress test which suggests clinically significant coronary arterial narrowing. In such a situation, coronary arteriography would seem reasonable and necessary to clarify the anatomic extent of the disease. The patient or family may be reluctant to follow such a recommendation. This reluctance has to be analyzed by the consultant with the patient. This reluctance is often based on a misconception of the nature and safety of the proposed procedure and may, in turn, be based on anecdotes of friends and relatives who have had poor results during the procedure or after the therapeutic recommendations emanating from the procedure. Those poor results may have little relevance to the patient's problem, occurring in individuals of different age groups or with different severity of disease or other complicating factors. A detailed discussion with the patient (and his or her family if the patient wishes) indicating the nature of the procedure, who performs it, how long it takes, its safety record, the importance of its results for significant therapeutic recommendations, and the consequences of not undergoing the procedure may be important in clarifying misperceptions. This may take time and further consideration by the patient; the decision usually does not have to be immediate.

Management problems may be quite complex, requiring recommendations for potentially toxic anti-arrhythmic agents, automatic defibrillators, cardiac transplantation, and treatment of complicated accompanying diseases.

A Hypothetical Case

A man in his forties or fifties is asymptomatic but has one or more risk factors for coronary artery disease: hypertension, hypercholesterolemia, cigarette smoking, diabetes, strong family history.

If hyperlipidemia is significant and diet does not solve the problem, one or more anticholesterol drugs may be advisable. In complex lipid disorders, the advice of a nutritionist or lipid specialist may be helpful.

The patient may develop angina, necessitating stress testing with a nuclear "tag" to clarify the anatomic state of the coronary arteries. A nuclear cardiologist is necessary to interpret the images. Further symptoms may warrant cardiac catheterization and coronary arteriography, for which subspecialists in these areas are necessary.

If intervention is necessary, a choice between surgery or some form of coronary angioplasty must be made, usually by the cardiologist and the referring internist. To carry out the recommendation, an "invasive cardiologist" or surgeon would be necessary, and additional subspecialty personnel may be needed in the course of the patient's illness.

If life-threatening ventricular arrhythmias develop, it may be necessary to study the cardiac electrical pathways and inducibility of significant arrhythmias as well as their response to therapeutic agents. A cardiac electrophysiologist should perform this study, and, in selected cases, provide direct therapy or prophylaxis by catheter ablation of the sites responsible for the arrhythmia.

The patient may develop a later cardiomyopathy with severe myocardial dysfunction despite adequate coronary blood flow. This may be refractory to conventional or even "experimental" therapeutic regimens and cause progressive disability. Cardiac transplantation may be advisable, requiring a therapeutic team consisting of transplant cardiologists, physiologists, social workers, and psychiatrists. This technique has become safer and more predictable, but in the waiting period for a donor heart mechanical circulatory support, which has its own experts, may be necessary. The technique of mechanical circulatory support is being used more frequently as a bridge to transplantation but probably will be used in the not-too-distant future as the definitive measure of circulatory support.

Assessing Therapies

The scope of the cardiologic consultant may be broad and may require substantial input from other physicians and scientists. A huge amount of knowledge is required in specialized areas. An ideal consultant must also have sufficient clinical experience and confidence in that experience so that he or she does not necessarily embrace the latest recommendations in the literature, but can assess the recommendations based to some extent on his or her clinical experience and analysis of the data in light of that (or opposite) experience. There are indeed fads or fashions in therapy that may not persist for long or are intuitively unsupportable. Experience and the confidence of experience are important in assessing these therapies for the patient. Patients are usually aware, from news broadcasts or publications, of the latest "advances" in cardiology as well as in other specialties. Reports in the lay literature are often fragmentary and overly laudatory or condemnatory and may not be prepared with a thorough or even a basic knowledge of the field. Such reports are often quotations from the conclusions of an article in medical literature or based on unsubstantiated claims that lack scientific rigor. The consultant must be prepared to pose and perhaps to answer the question: "It is new, but is it good?" In other contexts a therapy may be helpful in many patients but would not be appropriate or helpful in this particular patient.

The Art of Consultation

The art of the consultation was probably well known to the ancients. It requires gaining the confidence of the patient and close members of the family. This necessitates "quality" time: a careful history taken by the consultant without assuming that the history recounted by the referring physician is necessarily correct in all its nuances (for example, nature of chest pain, duration, precipitating factors).

A thorough physical examination not limited to the cardiovascular system is essential, as well as careful analysis of appropriate laboratory results and of specific cardiologic investigations: electrocardiogram, chest x-ray, Holter 24-hour electrocardiograms, echocardiogram, nuclear cardiologic findings, and anatomic and physiologic data obtained at cardiac catheterization.

As in most subspecialties, cardiologic problems often have a complex setting associated with other disease processes important for therapeutic recommendations. A competent subspecialty consultant must have good, broad-based internal medicine training and instincts, so that what is cardiologic and what is not can be recognized and treated appropriately.

Time must be spent in conversing with the patient and family members so that their intellect and medical sophistication can be assessed and explanations given appropriate to their level of understanding. It is important to have informal and free-ranging discussions covering the patient's occupation, lifestyle, aspirations, family constellation, and apprehensions while obtaining the specific facts in the history, because these factors are important in formulating a specific recommendation appropriate to that patient. The presence of the spouse or a close friend or relative who can expand the history and explain the patient's lifestyle is often helpful because the patient may distort, underexpress, or overexpress his or her symptoms and limitations.

After arriving at a recommendation, the consultant must be certain that the patient understands the facts involved and the nature of the advice. Too often, we physicians engage in "medspeak," jargon known to us but either unknown to or misinterpreted by the patient. Many highly intelligent, accomplished lay people often have rudimentary or erroneous perceptions of the anatomy and physiology basic to understanding the disease and its therapy; equally surprising is that the physician often does not appreciate these misconceptions and cannot or does not communicate effectively with the patient.

The art of the consultation must also include consideration of the sometimes delicate relationship between the patient and the family, the referring physician, and prior consultants in the same specialty and other specialties. The consultant's unquestionable primary duty is to the patient who requests his or her services, and advice must be given in light of this precept. Opinions differing from other consultants or treating physicians are frequent, and in my experience these dissents are often offered by highly qualified physicians. Regrettably, derogatory or gratuitously insulting remarks about the differing opinions and the physicians rendering them are sometimes made. These are purposeless and are most often unfair and rendered without full knowledge of the facts and circumstances known to the other physicians. The referring physician is most often the one who knows the patient best and must be consulted closely on recommendations to the patient. It is appropriate, of course, to return the patient to the care of the referring physician. Some complicating medical circumstances may require a degree of continuing care by both the referring physician and consultant. The patient may insist on concomitant care, sometimes for psychological reasons, and these cases have to be considered individually.

The desirable qualities of the clinical consultant include a solid clinical background in general internal medicine, highly specialized knowledge within cardiology, good knowledge of recent advances and trends in the field, and sufficient experience to have seen previous developments and have a firm idea of their history.

As in all physicians, the human qualities of compassion, empathy, and understanding and appreciation of psychologic factors are important. Frankness and good communication skills are equally important. A good sense of humor can lighten the load on the patient and the consultant and, where appropriate, an optimistic attitude can be helpful to the patient and family.

Caveats for the Future

In analyzing the amazing developments in cardiologic investigation and care during my own career, I believe that recent recommendations for the training of future cardiologists, some by government officials and some endorsed by medical educators, require some critique. There has been government initiative to ensure that at least 50% of our trainees enter primary care. Endorsements of this position have been rendered by, among others, The Council on Graduate Medical Education and The Association of American Medical Colleges. The impetus for this suggestion is the reduction of health care costs. This it might do as would other measures that would reduce the quality of health care. The latter should be our concern as cardiologists and as leaders of an important subspecialty.

I have not been able to determine why the figure of 50%, taken from projected needs of managed care programs, has been promulgated as a generally desirable applicable goal. Indeed, I do not believe it has a credible basis in firm knowledge of the number of specialists required in the future; that is truly an unknown (1). Although programs that lack quality should not exist and there may be some reduction of total physicians trained in the future, many advocates of a substantial reduction in subspecialist trainees seem to have selected a palatable general figure without recourse to the same rigorous analysis that is applied in reviewing a scientific article or proposed research project.

We do have people in leadership positions who are quite eager for a seat at the table and who really may be followers and not leaders.

Enormous advances in the diagnosis and treatment of cardiovascular disease, as in other areas of medicine, have occurred over the span of a relatively few years. All have been accomplished by intensive and extensive subspecialty training and have involved clinical and research efforts of large numbers of subspecialists. Not too long ago there were thought to be "too many" cardiologists for the needs at that time. Such prognostications, expressed rather forcefully at the time, did not anticipate the explosion of knowledge in so many fields that are of importance today: cardiac surgery, advances in imaging, electrophysiology, molecular biology, and cardiac pharmacology. All of these have advanced because of a critical mass of talented trainees stimulated by clinical and basic scientific investigators. It is likely that attempts to mandate restrictions on training of sophisticated cardiovascular clinicians and researchers by arbitrary quotas will stifle future advances in unpredictable ways. Advances have depended on and in the future will depend on many viable programs with adequate numbers of trainees. We must be careful to nurture our young and not direct them into fields that do not spark their enthusiasm, and we must make it possible for those with interest and talent to obtain compatible training of excellent quality.

The public is not motivated solely to have the most economical medical care. Over the course of years I have seen many patients who required important decisions on care. I have never seen one who opted for the most economical care. They want "the best" and "the best" usually involves subspecialists. Indeed, there is little question that many if not most cardiologic problems are managed more knowledgeably and perhaps more expeditiously by trained cardiologists than by general internists or family practitioners. One example recently analyzed is knowledge of drug therapy for acute myocardial infarction (2). We should not advocate reduction in training of cardiologists and other subspecialists in the future according to some preconceived, politically palatable formula that may turn out to be terribly wrong.

In considering the impact of "managed care," which is predicted by many to be the dominant mode of medical care delivery in the future, I believe it is also a mistaken perception to think much of the public would or should accept significantly limited access to specialists and to the specialists of their choice. Limitation of access to specialty care and specialists' procedures is a prime concept of many managed care plans and helps to provide profit margins for stockholders and large salaries for executives. In some, physicians share in profits if less care is delivered and share in the losses if more care is delivered. Access to a specialist is often important for patient well-being and functioning as well as for medical management. The privilege of choosing the consultant is also important to many, if not most, patients. They often feel a firm connection to the specialist of their choice based on strong recommendations by their or their families experiences with the consultant. Although some patients desire the most economical plan, others would not and should not accept, "Hi, I'm George. I'll be your cardiologist today." Managed care programs should consider these facts carefully before limiting access to specialists or choosing a particular specialist without giving patients reasonable options to preserve their freedom of choice. If we do not provide for future needs and advances it is later than we think for quality care.

References

1. Cooper RA. Seeking a balanced physician work force for the 21st century. JAMA 1994; 272:680–686.
2. Ayanian JZ, Hauptman PJ, Guadagnoli E, et al. Knowledge and practices of generalists and specialist physicians regarding drug therapy for acute myocardial infarction. N Engl J Med 1994; 331:1136–1138.

Involuntary Confinement for Tuberculosis Control:

The Jewish View

Fred Rosner, M.D.

Abstract

Patients with active tuberculosis who are noncompliant with their prescribed medical treatment program pose a serious hazard to others as well as to themselves. If directly observed therapy cannot be implemented, involuntary confinement of such patients is ethically justified in both secular and Jewish law. This discussion of the Jewish view on this issue is drawn from classical, biblical, talmudic, and rabbinic sources.

During the early part of the twentieth century, before the era of antibiotics, patients with tuberculosis were often confined involuntarily to specialized hospitals and sanitaria in an effort to prevent the spread of the then deadly disease (1). Between 1950 and 1985, the incidence of tuberculosis declined in the United States due to effective control measures, the availability of numerous antibiotics capable of eradicating the responsible microorganism in the vast majority of patients, and a general improvement in living conditions (2).

Since 1985, the number of cases of tuberculosis in the United States has increased each year because of a variety of complex social, economic, and medical factors. These factors include the emergence of the human immunodeficiency virus (HIV) and the acquired immunodeficiency syndrome (AIDS); large numbers of immigrants from countries where tuberculosis rates are high; the overuse of antibiotics; adverse social circumstances including poverty, homelessness, and substance abuse; and deterioration of the health care infrastructure (3).

Drug-resistant strains of tuberculosis have emerged. In New York City, a dramatic and perhaps alarming increase in the prevalence of multidrug-resistant tuberculosis occurred between 1985 and 1993 (4). High rates of treatment failure and relapse jeopardize patients' health and welfare and allow them to continue to transmit new infections. Thus, the acquisition of drug resistance imperils not only these patients, but those whom they subsequently infect. These conclusions have led some authorities to recommend directly observed treatment of tuberculosis (5) or even involuntary confinement for treatment.

An Institute of Medicine panel concluded that the two primary factors in the development of multidrug-resistant tuberculosis are patients' noncompliance with the 6- to 12-month regimen of antibiotic therapy required to treat tuberculosis and the prescription by physicians of ineffective treatment regimens (6). Because of increasing numbers of cases of multidrug-resistant tuberculosis, this disease has entered a dangerous new era and requires interventions to improve patient adherence to treatment regimens. Such interventions might include financial incentives such as money, bus tokens, or food; social rewards; educational interventions; directly ob-

---

From the Department of Medicine, Mount Sinai Services at Queens Hospital Center and Professor of Medicine, Mount Sinai School of Medicine, Jamaica, New York. Address reprint requests to Fred Rosner, M.D., Queens Hospital Center, 82-68—164 Street, Jamaica, NY 11432.

served therapy; and comprehensive services including total health care, shelter, job skills training, etc. (3).

Noncompliance is determined by a variety of social and personal factors. It is difficult even for healthy, intelligent, and highly educated people to take a combination of medications, sometimes including unpleasant and even painful injections, for 6 months to a year or even longer. The task is even more formidable for a person who is homeless and ill, who often has other major medical and social problems such as AIDS or drug or alcohol dependency, and who may have no family or support network.

The American Thoracic Society, the American Academy of Pediatrics, the Centers for Disease Control and Prevention, and the Infectious Disease Society of America point out that noncompliance can lead to treatment failure, drug resistance, continuing transmission of infection, increasing disability, and death. These learned societies state that if all other compliance-enhancing methods are not available or fail, "Consideration should be given to confining these patients to a hospital or other institution for treatment" (7). Such involuntary confinement would be used as a temporary measure until effective compliance-enhancing policies including incentives, social support, education, counseling, drug treatment, housing, and employment programs could be implemented.

Public health issues in the United States are mostly dealt with at the state level. A recently published survey of tuberculosis statutes in all 50 states contains proposals for reform (8). New York City public health officials are now confining uncooperative tuberculosis patients, against their will, for the duration of therapy. Until recently, the health code in New York City only allowed for the detention of patients until they were no longer infectious.

There is ample legal precedent for states (and cities) to set aside personal liberty and privacy when the public health and safety is at risk. Childhood immunizations against various infectious diseases are required in most states to protect health, even against the objections of rare individuals who claim such laws violate their constitutional right to privacy. Premarital serologic testing for syphilis was required until recently by law in most states to protect the public health. The history of confinement policies for other contagious diseases is summarized elsewhere (9) and is beyond the scope of this essay.

Involuntary confinement at home, in a hospital, or in a special facility of a patient with active tuberculosis who is a danger to others has considerable legal authority in the United States. However, ethicist Annas (10) points out that the burden of involuntary confinement will fall most heavily on the homeless and those who live in crowded, inadequate housing. He also states that the rationale for involuntary confinement is danger to others, based on the contagiousness of the patient's disease; therefore, patients have a right to be released when they are no longer infectious. However, if they then fail to comply properly with the continued requirement for antituberculous therapy, they may convert their previously sensitive disease to drug-resistant tuberculosis and thus become an even greater danger to others. The existing state laws that require release of patients when they are no longer infectious may have to be changed to permit or require confinement until cure is achieved. Treatment of tuberculosis must be individualized and the least restrictive and intrusive public health intervention should be used. Thus concludes Annas, "Directly observed therapy remains clearly preferable to involuntary confinement, however, and diligent and imaginative efforts to deliver therapy on an outpatient basis should be made before involuntary confinement is contemplated" (10).

This essay describes the Jewish view on involuntary confinement as found in classical biblical, talmudic, and rabbinic sources. The discussion focuses on imprisonment as a punishment for crime or as a measure for the public welfare or safety.

Involuntary Confinement in Judaism

Imprisonment as a form of punishment is not part of the judicial system in Judaism with rare exceptions. A person found guilty of a sin for which the punishment is 39 lashes was flogged as punishment, which it was hoped will lead to repentance for the sin (11).

A person who killed another human being inadvertently went into exile to a city of refuge for involuntary confinement. These cities belonged to the Levites and were specifically designated to afford asylum for an unintentional slayer (12). Exile to a city of refuge in Judaism was viewed as a matter of punishment as well as rehabilitation.

The unintentional slayer was not completely guilty of gross negligence or criminally negligent homicide because he had no intent to kill anyone. However, some forethought or precautions might have saved the victim's life. The Levites who owned the cities of refuge were primarily teachers who could sensitize the unintentional slayer exiled there about the high degree of respect we must have for the infinite value of human life and the lengths to which we must go to protect and preserve it.

The Jewish correctional system of sending inadvertent slayers to a rehabilitation city of refuge rather than a prison was to resensitize such people who showed callous disregard for human life and its supreme value. The punishment of exile was also to bring forgiveness.

Thus, the philosophy of imprisonment or incarceration as both punishment and re-education is an example of the invasion of privacy in the face of overriding societal needs. The analogy to tuberculosis, however, is not complete because patients with this disease have committed no crime. Patients noncompliant with their treatment, however, pose a significant risk to society and to themselves, so that involuntary confinement for treatment may be justified if no other option, such as directly observed treatment, is available.

The use of involuntary confinement of individuals to protect the public health and welfare and/or the use of quarantining for infectious and contagious diseases are discussed in the next section of this essay.

Societal Obligations Versus Individual Rights

The constitutionally guaranteed right to privacy in the United States requires physicians and other health care providers to obtain informed consent from patients before administering any therapy. Individual rights, some of which are legal and others moral, include the right to refuse treatment, the right to health care, the right to die, the right to be (or not be) resuscitated, the right to abortion, and many more. Although much of the modern secular ethical system is based on rights, Judaism is an ethical system based on duties and responsibilities. Rabbi Jakobovits eloquently articulates this Jewish view as follows:

> Now in Judaism we know of no intrinsic rights. Indeed, there is no word for rights in the very language of the Hebrew Bible and of the classic sources of Jewish law. In the moral vocabulary of the Jewish discipline of life we speak of human duties, not of human rights, of obligations, not entitlement. The Decalogue is a list of Ten Commandments not a Bill of Human Rights. In the charity legislation of the Bible, for instance, it is the rich man who is commanded to support the poor, not the poor man who has the right to demand support from the rich.
>
> In Jewish law a doctor is obligated to come to the rescue of his stricken fellow-man and to perform any operation he considers essential for the life of the patient, even if the patient refuses his consent or prefers to die. Once again, the emphasis is on the physician's responsibility to heal, to offer service, more than on the patient's right to be treated (13).

There is ample precedent in Judaism for setting aside individual rights, such as they are, to protect the health and welfare of society. The entire thirteenth chapter of the book of Leviticus deals with an illness called *tzaraat,* thought by most writers to be leprosy (14). This illness was thought to be so dangerous that an individual afflicted with it was completely isolated from the company of healthy people. Acute and chronic forms of this skin disease existed. In Jewish law the leper was confined by examination and statement of the priest (analog of the physician). Biblical leprosy also has spiritual and moral aspects to it. In fact some writers consider the entire chapter on leprosy to be an allegory, as if Scripture speaks of "leprosy of the soul." The involuntary confinement or quarantining of affected individuals was imposed, however, whether the disease was physical or spiritual. A specific instance of quarantine for leprosy is the case of Miriam, the sister of Moses, who suddenly became leprous as snow when she spoke slanderously about her brother. She was quarantined but was healed after 7 days as a result of her brother's prayer (15). The prophet Elisha's servant Gehazi was similarly treated (16). It may be that "leprosy" marks the sinner and separation from the community is the punishment. Does moral culpability with resultant involuntary confinement describe noncompliant tuberculosis patients? Perhaps. It is also possible that social circumstances may result in their noncompliance thus at least partially absolving them of moral culpability.

Societal obligations to protect the health and welfare of its citizens even at the cost of ignoring individual rights are exemplified in Judaism in numerous contexts. For example, the Talmud asserts that captives should not be ransomed for more than their value, to prevent abuses (17) either so that captors should not be encouraged to seize more captives and make excessive ransom demands or so as not to impoverish the community. If society has to pay large sums of money to ransom kidnap victims, that money would not be available for other pressing societal needs. Society must serve all its people and must occasionally set aside the right of an individual to be ransomed in order to serve the greater good of serving all the people. Although individuals must be taken into account in societal decisions, there is more to society than the individuals in it (18).

Individuals should also be concerned about the future and the needs of subsequent generations. This view is most beautifully depicted in the talmudic story of the righteous man Choni the Circle-Drawer, who was journeying on a road and saw a man planting a carob tree. He asked him: "How long does it take for this tree to bear fruit?" The man replied: "Seventy years." Choni then further asked him: "Are you certain that you will live another seventy years?" The man replied: "I found ready-grown carob trees in the world: as my forefathers planted these for me, so I too plant these for my children" (19).

Judaism also requires the negation of person's right to smoke in the presence of others to protect the latter from the harmful effects of passive smoking. In 1980, Rabbi Moshe Feinstein (20) clearly and definitively prohibited the smoker from smoking in a place where others would be subjected to passive smoke. In 1982, Rabbi Yechezkel Grober (21) wrote that smokers should be firmly but diplomatically asked to stop smoking in the presence of others so as not to bother either healthy people or sick patients. Rabbi Eliezer Y. Waldenberg (22) also ruled that a smoker can be prevented from doing so in public if any nonsmoker is bothered by the smoke or might become ill as a result of the passive smoking.

A final example of setting aside an individual's rights for the sake of avoiding harm to a third innocent party is the Judaic concept of *rodef* or pursuit. If A is pursuing B with the intent of killing B, C must prevent the pursuer from carrying out that act. If lesser measures fail, C should kill A to prevent A from killing B. The rule of the pursuer, according to Maimonides and others, is based on a biblical passage where a woman is pursuing a man with the intent of killing him (23). The Bible rules that she must be stopped because one is biblically prohibited from standing idly by without attempting to rescue one's fellow man whose life is threatened (24). Based on these biblical passages, the Talmud states that he that pursues his fellow man with intent to kill may himself be killed first to prevent him from doing so (25). It is one's duty to disable or even take the life of the assailant to protect the life of one's fellow man. The Talmud gives the example of a thief who breaks into a man's house and who, if discovered, might kill the homeowner (26). Killing of the thief as an act of self-defense is permitted because the thief is considered as a potential murderer or pursuer endangering the life of the homeowner. A noncompliant patient with tuberculosis who constitutes a serious risk to other people's health and lives may therefore be stopped by involuntary incarceration to prevent him from infecting others.

Summary and Conclusions

Man cannot survive in society without a system of laws and ethical principles and without a government or other protective authority. In modern secular society, individual rights and autonomy reign supreme. However, government has to be concerned not only for individuals but also for the public welfare. Therefore, individual rights must sometimes give way to the greater good of the protection of society as a whole.

The rights of an individual patient with tuberculosis who is noncompliant with his therapeutic regimen and thereby endangers others around him must be set aside for the public health and welfare. If all other methods to achieve compliance fail, quarantine or isolation of the patient may be used, although Judaism would urge the least restrictive practice such as directly observed treatment. Many states in this country specify the power to isolate or quarantine in their home individuals who endanger the public health. In many states the powers are very broad, authorizing isolation in a jail, hospital, or other facility. The majority of states allow isolation or quarantine through a public health order without specific due process (8). Judaism supports this approach.

Judaism considers that a human life has infinite value. Therefore, physicians and other health care providers are obliged to heal the sick and prolong life. Patients are also duty bound to seek healing from a physician when they are ill and not rely on divine intervention or faith healing. Patients are charged with preserving their health and restoring it when ailing in order to serve the Lord in a state of good health (27). A patient with active tuberculosis is not only a danger to society but also to himself. He is obligated in Jewish law to accept therapy that is effective and potentially curative. If he refuses it, he should be coerced to accept it.

The restoration of health, the induction of remission from illness, and the saving of lives are not only individual obligations, incumbent upon physicians, patients, and others who are able to do so, but also communal or societal obligations. Society is obligated to protect its citizens, to care for the health and other needs of its people, and to take appropriate action to ensure that these goals are met. Sometimes, to accomplish these goals, society must intrude on and even negate individual rights such as the right of a patient with tuberculosis to refuse treatment, thereby endangering other people and himself. Direct observed treatment is the preferred method (10), but, if that fails, involuntary confinement of the patient is ethically justified.

References

1. Ayvazian LF. History of tuberculosis. *In:* Reichman LB, Hershfield ES, eds. Tuberculosis. A Comprehensive International Approach. New York: Marcel Dekker, 1993: 1–20.
2. American Thoracic Society. Control of tuberculosis in the United States. Am Rev Respir Dis 1992; 146s:1623–1633.
3. Sumartojo E. When tuberculosis treatment fails. A social and behavioral account of patient adherence. Am Rev Respir Dis 1993; 147:1311–1320.
4. Frieden TR, Sterling T, Pablos-Mendez A, et al. The emergence of drug resistant tuberculosis in New York City. N Engl J Med 1993; 328:521–526.
5. Iseman MD, Cohn DL, Sbarbaro JA. Directly observed treatment of tuberculosis—we can't afford not to try it. N Engl J Med 1993; 328:576–578.
6. Lederberg J, Shope RE, Oaks SC, eds. Emerging infections: microbial threats to health in the United States. Washington, DC: National Academy Press, 1992.
7. Bass J, Farer L, Hopewell P, et al. Control of tuberculosis in the United States. Am Rev Respir Dis 1992; 146: 1623–1633.
8. Gostin L. Controlling the resurgent tuberculosis epidemic: a 50 state survey of TB statutes and proposals for reform. JAMA 1993; 269:255–261.
9. Merritt DJ. The constitutional balance between health and liberty. Hastings Center Rep 1986; 16(6):S2–S10.
10. Annas GY. Control of tuberculosis—the law and the public's health. N Engl J Med 1993; 328:585–588.
11. Moses Maimonides' Mishneh Torah, Sanhedrin 18: 1 ff.
12. Moses Maimonides' Mishneh Torah, Rotzeach 5: 1 ff.
13. Jakobovits I. The timely and the timeless; Jews, Judaism and society in a storm-tossed decade. New York: Bloch, 1989: 128.
14. Preuss J. Biblical and Talmudic medicine (F. Rosner, Trans.) New York: Sanhedrin Press, 1978: 323–353.
15. Numbers 12:11–13.
16. 2 Kings 5:27.
17. Tractate Gittin 2:6.
18. Tendler MD. Rabbinic comment on triage of resources. J Mt Sinai Hosp 1984; 51:106–109.
19. Tractate Taanit 23a.
20. Feinstein M. Cigarette smoking that bothers others in the study hall. Responsa Iggrot Moshe, Choshen Mishpat Part 2 #18.
21. Grober Y. Is it permissible to smoke in a group where some people object to the smoke. Am Hatorah Vol. 2, No. 3, 5742 (1982): 92–102.
22. Waldenberg EY. Responsum on smoking in halachah. Assia No. 35, Vol. 9 #3, Feb. 1983, pp. 10–15.
23. Deuteronomy 25:11–12.
24. Leviticus 19:16.
25. Tractate Sanhedrin 72b–73a.
26. Sanhedrin 72b.
27. Rosner F. Modern medicine and Jewish ethics. 2nd ed. Hoboken, NJ and New York, NY: Ktav and Yeshiva University Press, 1991: 519.

The Need for More Primary Care Physicians to Contain Costs and Improve Access:

Myths Versus Realities

Barry Stimmel, M.D.

Abstract

It is commonly believed that there is a surplus of specialists in this country as well as a deficit of primary care physicians. To remedy this, pressure has been exerted on medical schools to change their curriculum and to specifically assure that at least 50% of all graduates will enter primary care careers. The data, however, suggest that although we have an excess of specialists, there is no deficit of primary care physicians. Changes that are occurring in the medical marketplace have already resulted in more residents entering primary care fields independent of curricular changes. The training of both medical students and residents, however, must be revised to prepare physicians for the practice of medicine in today's economy. These changes should be fueled by the desire to produce as well trained a physician as possible, rather than a responsibility to graduate specific numbers of primary care physicians.

It is now commonly accepted that we are facing a crisis in the provision of primary care in no small part due to our having too many specialists and far too few generalist physicians. In an attempt to remedy this deficiency, the Robert Wood Johnson Foundation has made available $32 million to encourage medical schools to select and train students who will become family practitioners, general internists, and pediatricians (1). The Association of American Medical Colleges (AAMC), the Council on Graduate Medical Education, the Association of Professors of Medicine, the American College of Physicians, the American Academy of Family Practice, the American Society of Internal Medicine, and the Bureau of Health Professions of the Health Resources and Service Administration, not to mention numerous commissions and foundations, all agree that 50% of students graduating from medical schools in this country and 50% of physicians leaving residency training should become "generalists" or primary care physicians (2). Buoyed by these declarations and concerned by an obvious lack of access to medical care by a large number of our population and a "health care industry" consuming 14% of our gross national product, legislative remedies are rapidly appearing. These attempt both to coerce and to entice students, schools, and physicians to fill the "ranks of the generalists," as well as schools to revise their curricula to ensure that this objective be met. Similarly, the public is being provided with solutions to this problem, the implications of which remain poorly defined.

Yet, despite the obvious and often repeated observation that proportionally fewer students graduating from U.S. medical schools are choosing careers in primary care, little attention has been focused on whether the real problem is the

---

From the Mount Sinai School of Medicine of The City University of New York, New York, New York. Address reprint requests to Dr. Barry Stimmel, Dean for Graduate Medical Education, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1193, New York, NY 10029.

numbers of primary care physicians, the adequate utilization of existing physicians' services, or the inadequate access of many to medical care. This paper will review existing myths and realities associated with the provision of primary care as well as the purported solutions to this problem. In this context primary care providers will be defined as family practitioners, general pediatricians, and general internists. Although many believe that obstetricians are also primary care providers, it is unclear to what extent this occurs.

The Need for Greater Numbers of Primary Care Physicians

The recognition of a need for more generalists is a relatively recent phenomenon. In the 1980s, the Graduate Medical Education National Advisory Committee (GMENAC) predicted a future balance or slight surplus of primary care physicians (3). A revised GMENAC estimate for 1990 listed an excess of 5% each for osteopathic family practice, general family practice, and general internal medicine and a 15% surplus for general pediatrics, for a total surplus of 12,750 generalists (3, 4).

Despite the current belief that we are losing primary care physicians, the primary care specialties have, in fact, remained relatively stable over the past two decades with respect to total physician population. In 1970, 40.9% of physicians were in primary care specialties, as compared with 39.6% in 1990 (5). In terms of absolute numbers, however, the number of physicians engaged in primary care increased remarkably, from 136,637 in 1970 to 243,419 in 1990, a growth slightly in excess of 78%. Even if one looks in terms of absolute numbers, with respect to general practice or family practice, the number of these physicians was 21% greater in 1990 than in 1970. The number of all primary care physicians per 100,000 population increased from 69.6 physicians in 1980 to 115 physicians in 1990, an increase of 65%. In addition, these data ignore the considerable number of physicians who, although having received advanced training in a specialty of medicine, nonetheless, due to their interests or to the marketplace, actually spend a considerable proportion of their time engaged in primary care. The argument advanced that a person who has had two extra years of training as a subspecialist cannot be a fine primary care physician is flawed. This person has had the identical general medical training as those who practice only general internal medicine. In many programs fellows in a specialty may spend time in general medical clinics or emergency rooms as part of their house staff requirements or as "moonlighters." In fact, the existence of primary care internal medicine as a distinct residency is a relatively new phenomenon and represents a small portion of internal medicine positions (10% in 1994), with the major difference between these and general medical residencies being the time spent in an ambulatory care setting.

The current recommendation that 50% of all physicians in this country be in primary care was based on the experiences of other industrialized nations (6). Yet, this is a jump in logic as one is relying on percentages, not physician/population ratios. To place this in an international perspective, we are, in fact, in a much better position than many countries known for their focus on primary care. In Great Britain in 1989, there were approximately 69.7 general practitioners per 100,000 population (7). In Britain, only physicians in general practice provide primary care. Indeed, they have been providing primary care to proportionally more people with considerably fewer physicians than are currently engaged in these specialties in this country.

If one views Canada, where 50% of physicians are in primary care, with one physician per 1,149 persons for a population of 26,835,500, there are approximately 111 primary care physicians per 100,000, slightly less than that seen in this country (8). In Ontario, there are 90 general physicians per 100,000. Yet, in this province the government has mandated decreasing medical school enrollment. It is also often forgotten that one of the main reasons why 50% of Canadian graduates choose primary care may not be because of a greater interest but rather that the number of graduates closely approximates the number of residency positions, 50% of which are in primary care fields. Were this the setting in the United States, we immediately would have half of our graduates choosing primary care in the absence of either curricular reform or financial incentives.

It is difficult to determine with any precision how many primary care physicians are needed to provide adequate medical care. As noted by Stoddart and Barer in a 1985 World Health Organization survey of thirteen countries asked to assess the adequacy of their physician supply, there was little relationship between the number of physicians per 100,000 population (which varied from 8 to 228 physicians) and the perception of a physician shortage (9). However, although the United States has proportionately fewer physicians in primary care, we nonetheless seem to compare quite favorably with respect to primary care physician/population ratios with other countries. Indeed, the latest American Medical Association data revealed that if one views only those physicians in active clinical practice, physicians in the primary care specialties account for 45% of all physicians (5). It should be noted that these data have been criticized by both "camps." Those seeing a need for primary care physicians correctly point out that this data base only distinguishes the specialties of cardiology, pulmonary diseases, gastroenterology, and allergy, lumping together all other specialties within general internal medicine. Others, however, also correctly note that physicians in these "separated" specialties may spend a considerable proportion of their time providing primary care to their patients.

Further, despite all predictions, with the rapid growth of managed care, it is really impossible to accurately assess whether a shortage or surplus of physicians will exist in the year 2000 as staffing patterns in these groups result in 20% fewer physicians for children and 50% fewer physicians for adults. However, chronic illnesses, such as AIDS, the increasing number of women physicians who justifiably take time from their practice to raise children, and a potential shorter work week for physicians on fixed salaries may obviate any physician shortage (10). It is, however, quite true that not only proportionately, but also in absolute numbers, we have far more specialists than seen in other developed nations. This is costly and is an issue that should and is being addressed based on the level and quality of medical care we wish to receive. It is, however, not germane to a need for primary care physicians.

Increasing Vacancies in Primary Care Residency Positions

There is no question that the proportion of family practice residencies filled through the National Resident Matching Plan (NRMP) has diminished. In 1984, 85% of positions offered in family practice were filled as compared with 68% in 1992. However, in 1994, 83% of all positions filled. Comparable figures for medicine residencies (88% vs 82% vs 89%) and pediatric residencies (88% vs 82% vs 92%) are far better (11). It is also true, but less often realized, that although the proportion of total primary care positions offered in the NRMP has remained fairly constant between 1981 and 1994, in fact, the number of positions in these fields has increased by approximately 1,800, without any evidence to justify such increases in view of GMENAC's original survey on manpower needs. The call to have 50% of all first-year residency positions in primary care may appear revolutionary, yet this is clearly not the case. If one were to eliminate all preliminary and transitional positions from those offered in the NRMP, under the assumption that one must enter a second-year position in another specialty, then 48% of all NRMP positions are already in primary care. It is recognized that at present a substantial proportion of those completing internal medicine residencies (estimated to be as high as 60%) and, to a lesser extent, those completing pediatric residencies go on to specialize. This, however, can be simply addressed by decreasing the total number of specialty fellowships without holding either schools or students to fixed quotas. Indeed, anecdotal evidence suggests that in fact many specialties in internal medicine are considering a substantial downsizing of their fellowship programs. Further, the accrediting agency for surgical residencies has recently announced its intention to diminish all preliminary surgery positions to the numbers of positions available for advanced surgical placement in orthopedics, otolaryngology, neurosurgery, and urology.

In addition, the current increasing, if not explosive, interest of U.S. citizens in attending medical school suggests that the problem of filling primary care residency positions will soon become moot. In 1993, there were 42,808 applications for the Class of 1997, breaking the previous high seen in 1974, when 42,303 students applied (12). During the 1994–1995 academic year, the number of applicants exceeded all previous records. It is fairly safe to assume that as the number of applicants to U.S. medical schools increases, so will the number of U.S. citizens accepted at foreign medical schools. In 1984, 3,347 U.S. citizens in foreign medical schools applied to the NRMP. This represented 11.6% of total applicants. In 1992, this number had decreased by 69% to 1033 and represented only 4.1% of the applicant pool.

Although the number of active applicants (those who actually complete the matching process) in each category is considerably less (**Table 1**), it is not unreasonable to assume that increasing numbers of unsuccessful applicants to U.S. medical schools will, in fact, pursue their medical education outside of this country and, after successfully obtaining their ECFMG certification, return to residency training and ultimately practice in the United States. If one views the number of U.S. foreign medical graduates (USFMGs) applying to the NRMP as a proportion of the unsuc-

TABLE 1
Selected Applicant Groups to Matching Program

	1981	1984	1991	1992	1995
U.S. seniors	13,705	14,741	13,943	14,030	14,621
U.S. foreign medical graduates	536	1,695	601	617	735
Alien foreign medical graduates	1,127	4,212	2,259	3,196	5,575
Fifth pathway	350	339	59	32	16
U.S. physicians	181	768	903	901	1,127
Total active	16,104	22,052	18,450	19,519	22,936

cessful applicants to U.S. medical schools 4 years earlier, then between 1980 and 1988, this proportion varied from 10.2% to 15.4%. Applying these percentages to the applicant pool of 43,000 students this year, as well as to a projected applicant population of close to 45,000 next year, then in 1997 one would predict between 2643 and 3991 USFMGs entering the NRMP, with between 4377 and 6609 USFMGs entering the following year.

The number of non–U.S. citizen international medical graduates (IMGs) entering the NRMP has also been increasing over the past several years, from 1127 active applicants in 1981 to 5574 in 1994. In 1994, there were 0.9 positions per candidate of all first-year positions offered in the NRMP. Although it is true that IMGs tend to prefer specialty training more than graduates of U.S. medical schools (USMGs), in fact, as demonstrated by matching plan data, they tend to fill those positions considered least desirable to USMGs, regardless of whether in primary care or specialty programs. As the total number of available positions in the NRMP will in all probability diminish with the emphasis on specialty positions, virtually all of the positions in the primary care fields will soon become filled. If one allows only a 10% excess in first-year residency positions, as currently suggested, all available primary care residency positions will certainly be filled, assuming it will be the specialty positions that will diminish in number.

Diminishing Interest of U.S. Graduates in Primary Care

It is impossible to escape the conclusion that over the past several years fewer USMGs have been interested in practicing primary care. In the 1994 March, USMGs filled 67% of family practice positions and 54% of internal medicine positions, as compared to 75% and 73%, respectively, in 1984. Although family practice seems to have been more affected than internal medicine, if the number of internal medicine residents proceeding to specialty training is considered, this is probably not the case (13). In addition, the number of U.S. physicians who have already had at least 1 year of graduate training and wish to enter a different specialty has increased from 181 in 1982 to a high of 1045 in 1990, leveling off over the past 2 years at approximately 900 (11). Although data are not readily available as to the reasons for the desire to change careers, the nonmatch rate for this group of 51% suggests the movement is not from a specialty to a generalist career.

The reasons for the diminishing interest of U.S. graduates in entering primary care have been discussed in detail by a variety of authors and organizations (14–16). All of these factors (Table 2) to some extent play a part in this lack of enthusiasm. All must be addressed if one is to change the perception of our students about primary care medicine. Although each of these issues can be addressed quite easily "philosophically," pragmatically some are more amenable to solution than others. Of primary importance, however, and the factor most often ignored, is the existing compensation for primary care physicians as compared with specialists. Levinsky (17), in reviewing the evidence, suggests a direct linear relationship between the average income of physicians in a specialty and the success of residency programs in filling positions with U.S. graduates. A 1% increase in annual income is believed to increase the number of those entering that specialty by 0.5%. Levinsky suggests that were this relationship to be accurate, then the recently adopted Medicare relative value scales would increase the entry of U.S. graduates into primary care by less than 6%.

TABLE 2
Reasons Cited for U.S. Medical Graduates Not Entering Primary Care

Insufficient undergraduate role models
Insufficient curricular time in primary care in the medical schools
Lifestyle
Marked discordancy in incomes of primary care vs specialty medicine
Increasing financial indebtedness
Increasing bureaucracy in primary care practice
Diminishing enthusiasm expressed by primary care physicians
Diminishing prestige associated with primary care
Increasing interest in technology of modern medicine
Inadequate training programs

In fact, indirect evidence suggests that, independent of curricular changes, this is already occurring. As the salary paid by managed care groups has increased, that paid to specialists has decreased correspondingly. As a result, in many parts of the country primary care physicians leaving residencies are making more money than their counterparts finishing specialty training, with some specialists, notably anesthesiologists, having great difficulty finding jobs. It is no coincidence that over the past 2 years the number of USMGs entering anesthesiology has decreased precipitously, with the proportion entering primary care fields increasing. In the 1995 Match, 87% of family practice positions filled compared with 65% in 1991, with the percentage of positions filled by USMGs increasing from 56% to 71%. General internal medicine positions filled to 92% as compared to 81% in 1991, with the filling of pediatric positions increasing from 79% to 96% during this time.

Unfortunately, all too often the approach taken is coercive. This frequently is not only ineffective, but even counterproductive. Two such examples come quickly to mind. Medical schools are now being called upon to assume responsibility for graduating 50% of their students committed to primary care—even at the expense of altering their criteria for admission. This is a perfect example of the ends justifying the means. Considering the tremendous increase in the number of applicants to medical school, there is no doubt that more students are already beginning to express an interest in primary care, regardless of their ultimate career goals. Because so few students really know what their interests are on entry into medical school, forced selection at an unreasonable time encourages dishonesty. This perception is not lost on the students and will remain throughout their careers. The obligation of a medical school must be to provide the best possible education to permit students to practice medicine in the 21st century in accordance with public needs and modern science. There is no question, as discussed below, that schools can and must do better and, by including primary care education in the required curriculum, may influence students' ultimate career choice. However, insisting that a school assume responsibility for a proportion of students actually entering a field is beyond the pale.

The second example of a good intention going bad is the recent changes in the Health Professions Student Loan (HPSL) program through reauthorization of the Higher Education Act (Title IV) and the Health Professions Act (Title VII). These changes, implemented virtually without any public discussion, converted an excellent, need-based loan program into a Primary Care Loan (PCL) program requiring a commitment in the first year of medical school from all borrowers to complete a primary care residency and practice primary care until all PCLs are repaid in full. Failure to meet the commitment results in the conversions of the PCL from a 5% interest rate to 12%, recomputed and compounded annually. Furthermore, all schools must monitor the borrower's compliance as well as ensure that 50% of its graduates will ultimately enter primary care. This must be done by increasing the percentage points of graduates entering primary care by 5% each year until 50% is reached. The penalty for noncompliance on the part of the school requires a return of 10% of HPSL money in 1993/1994, 20% in 1994/1995, and 30% in all subsequent years. These requirements are remarkable as no new federal dollars are being put into the program. The program seeks to influence the career choices of 50% of students while providing minimal financial assistance for only 3%—a disproportionate effort to say the least. In addition, all of the money "feeding" the system is from repayments of graduates who have had these loans.

In an attempt to assist schools in submitting plans to meet these goals (which for many schools are impossible), the Department of Health and Human Services has helpfully offered advice, such as establishing special selection committees to prioritize candidates for the PCL program, adjusting memberships and policies of admissions committees, and altering mission statements of medical schools.

Even more important than the burden being placed on the schools, however, is the implication these changes place on social policy. Because the greatest proportion of the recipients of HPSLs are underrepresented minorities, this policy assumes that such students have a special obligation to practice primary care. At best this view is shortsighted and inappropriate. In fact, underrepresented minorities are most underrepresented in the specialties, not the primary care fields. Further, if the total pool of residency positions is relatively fixed, if the number of applicants increases (as will probably occur until the two are equal), or if there are more applicants than positions, movement of USMGs into primary care fields will result in the specialty fields being filled by international medical graduates. This will obviously do nothing for specialty practice but merely alter the distribution of income based on medical school attended.

More Primary Care Physicians Will Diminish the Costs of Medical Care

It is commonly accepted that we rank number one in the world in expenditures on health care spending, with an estimate that by 2030 we will spend almost one third of our gross national product on health care (18). It is true that physicians bear a responsibility for this cost. This is not through personal incomes, which, as demonstrated by Fuchs (18), if immediately cut by 20%, would reduce the health care bill by only 2%, but through generation of other health care expenses. It has been estimated that every physician added to the pool of practicing physicians generates an additional $800,000 in health care costs (19). The assumption is related to the beliefs that (1) much of these costs are due to specialty practice through increasing use of new technologies and (2) more primary care physicians would result in fewer specialists' costs. However, although evidence suggests that the first assumption is correct, available data are much less convincing with respect to the second.

Canada is most often touted as the country achieving the most success in containing costs of physician services. However, Hughes (20), examining physician services in the provinces between 1975 and 1987, found a steady increase in utilization of services, in some cases exceeding that of the United States. On the basis of physicians per 100,000 population, the percent increase in Quebec was 42.5%, as compared with 34.4% in the United States. Referral or consultation patterns by primary care physicians have also been shown to vary greatly in both Great Britain and the United States, with little correlation found between referral rates and quality of the referrals (21). Furthermore, Canadian health authorities appear to be rethinking their position with respect to the need for primary care physicians. The number of students admitted to Canadian medical schools will be decreased by 10%, with the University of Toronto having decreased its first-year enrollment by an even greater percentage. Reimbursements for new primary care physicians were proposed to be cut by 75% in Ontario, and although this proposal never came to fruition, many physicians feel that the real intent of the government is to have physician assistants provide primary care rather than physicians (22).

Of more importance, however, is the question that only a few have dared raise: What is wrong with spending 14% of our gross national product on medical care? As described by Seltzer (23), we do not compare the costs of other quality-of-life issues with other countries and determine we are spending too much. Our standard of living with respect to amenities, such as central air conditioning, large cars, and entertainment, exceeds that of most other industrial nations, yet few complain. Those countries that spend less on health care do so at a cost that has yet to be made clear to the public, and that cost will only be accentuated by embracing the concept of for-profit managed care.

Managed Competition as a Solution to Rising Health Care Costs

Recently, much emphasis has been placed on managed competition to diminish health care costs. Although by providing financial incentives for not making referrals, by changing fee-for-service to capitation, and by placing more management in the hands of businessmen rather than physicians may be accompanied by decreasing health care costs, issues of quality frequently arise as do those of conflict of interest (24, 25). In addition, the increasing bureaucratic costs needed to maintain such a system may well serve to diminish any real savings. However, based on past experience, it is fairly certain that the profits of the third-party carriers would be assured.

As documented by Himmelstein and Woolhandler (26–28), the administrative costs of third-party carriers not only play a significant role in contributing to health care costs but, in addition, far exceed those seen in single-payer systems. In 1989, third-party carriers received $18.7 billion for administrative costs and profits, with administrative overhead being 11.9% of premiums. This is compared with a 3.2% administrative overhead in Medicaid and Medicare and a 1.2% administrative overhead in Canada. These authors noted that Blue Cross and Blue Shield of Massachusetts employed 6682 workers to process bills for 2.7 million subscribers—a greater work force than in all of Canada's provincial health plan covering over 25 million people. More recent reports of the efficacy of Blue Cross and Blue Shield in New York do little to assure one that mismanagement and waste not only continue to exist but will become more prominent in a "managed care" society (29). If our efficiency could be brought in line with

Canada's, in 1987 up to $83 billion could have been saved. This might well be sufficient to provide health insurance for the millions of uninsured in this country without any changes in the system.

However, in the United States, the "cost management" industry is one of the fastest growing segments of the "health care business," with $7 billion in revenues estimated in 1993 (28). If one reviews a prime example of managed competition, the California Public Employees Retirement System (CALPERS), which purchases health insurance for 900,000 employees, cost containment has not seemed to be prominent. From 1982 to 1992, premiums for families exceeded the national average of 9.4% by up to 12.9% for its fee-for-service preferred provider organizations (PPOs) (30). This will undoubtedly change as competition among managed care groups has sharply increased. Indeed, CALPERS has already decreased its premiums in response to the marketplace.

Clearly, with increased administrative or overhead charges taken "off the top," managed competition is the panacea for controlling health care costs only when accompanied by a decrease in provision of services. It is whether this decrement consists of essential or excessive health care that will ultimately allow one to assess the acceptability of managed care for the public good (31–33).

However, as described by Davis (34), with the tightening up of fee-for-service practices, the decreasing rates of hospitalization, and the increasing use of ambulatory care settings for procedures usually requiring hospitalization, the differences between standard fee-for-service and managed care have narrowed considerably. For those enrolled in managed care other than staff and group model HMOs and programs based on the Kaiser model, the savings appear negligible and the yearly increases in cost of care comparable (34). If, in fact, many managed care systems do not seem to greatly reduce costs and do increase administrative overload, the only way the uninsured can be included without increasing spending is through diminishing cost-inefficient practices, such as the physician spending time with the individual patient or decreasing the availability of high-cost technical services. Even now, without adding the burden of the uninsured or the critically ill patient, satisfaction with HMOs is significantly less than that seen with solo or single specialty practices considering all aspects of medical care (35). Indeed, the mandate to constrain costs and increase access can only ultimately lead to widespread dissatisfaction on the part of the more than 85% of the population that currently is insured. Perhaps the ultimate example of this model can be found in Japan where a physician can see between 50 and 60 persons a day, rarely spending more than 5 minutes in each encounter.

The Role of the Medical Schools

There should be no doubt that there is a great need to both increase and enhance the teaching of primary care in the undergraduate curriculum. The raison d'etre for this, however, should not be the lack of primary care physicians, but rather the changing mix in patients seen in the hospital setting, as well as the need to provide high-quality education to students, regardless of the fields they will ultimately enter. Indeed, many academic family medicine departments have been able to accomplish this objective.

Since the publication of the Flexner Report in 1910, medical schools have established the inpatient service as the basic model for student education (36). This was appropriate as the general medical services of an institution offered a broad mix of cases, which could be evaluated by the student at a fairly leisurely pace under supervision, not only by attendings or preceptors but by an everpresent cadre of enthusiastic house staff. Attendings could schedule time away from their practices to meet with students, and the residents could continue individual student instruction on a daily basis. Quality control could be maintained more easily and the cost/benefit ratio was quite high. Unfortunately, the practice of modern medicine has changed all this.

Hospitals are increasingly becoming short-term stay facilities with the exception of the chronically or terminally ill patients. It has become both financially and logistically desirable to treat patients on an ambulatory, rather than an inpatient, basis. These changes mandate that clinical education be reviewed carefully and altered appropriately. Even here, however, thinking tends to be clouded. Schools are being encouraged to increase involvement of students in "preceptorships" by assigning them to individual physicians' offices, rather than developing high-quality ambulatory care facilities located either within the medical school setting or within health practice plans that specifically allocate staff time for teaching.

The preceptorial method was the prevailing situation when Abraham Flexner published his report concerning the quality of medical education in American medical schools. To quote Flexner:

> The backbone of the structure is the clinic in internal medicine. This central fact cannot in America be too strongly emphasized. The sufficiency of the school's clinical resources depends at bottom on its medical clinic; the value of its training depends on its systematic thoroughness with which it is in position to use an adequate supply of medical cases. The main point is that there is no substitute for a good clinic in internal medicine; the school sampled and found wanting there suffers from a fatal organic lesion (36).

It is necessary to emphasize immediately, so as not to be misunderstood, that patients may be receiving excellent medical care in private practices, and their physicians can and should serve as role models to students. Indeed, it is just these physicians who in many teaching programs serve exceptionally well as attendings and preceptors. Assigning students to private physicians' offices or to HMOs for their basic clinical education, however, most of necessity diminish the efficiency of the physician to see patients or diminish quality control even under the most optimal circumstances. Over a decade ago, several studies, attempting to assess the cost to HMOs of having students or residents receive their clinical education on site, found the annual cost per student to range from $16,900 to $21,480. In terms of 1990 dollars, the costs would be considerably higher (37–39).

This is not to suggest that students will not gain valuable experience by having practicing physicians as preceptors. However, there is a difference between having students take electives or selectives in physicians' offices and having them experience their basic clinical training at such sites. The first should be encouraged once their basic training has been received, with participating physicians receiving full recognition for their efforts. The second, although valuable, must be accompanied by careful monitoring of quality control as well as a realistic appraisal of the expenses that will be incurred.

Yet, despite the pressure being placed on schools to expand the teaching of primary care in their curriculum, little thought is being given to how truly high-quality primary care education can be funded. In the mid-1970s, when priority was given to expanding medical school enrollment, capitalization was provided by the federal government as well as many states. Despite the costs incurred with increasing class size being considerably less than that involved in providing quality primary care experiences, public assistance to accomplish this objective has been minimal. At present, tuition, which is already unacceptably high, accounts for only 16% of the funding for medical education costs, representing only 4% of alternative school revenues (40). In fact, total spending for both medical school and residency training is only 3% of all health care expenditures. If the public interest can best be served by the fabric of medical education and this alteration is exceptionally expensive, should not assistance be provided to allow it to be done correctly?

With provision of role models and increasing exposure to primary care experiences as part of the required curriculum, more students may consider becoming generalists. However, even with the best intentions and appropriate financial support, schools cannot significantly influence the career choices of their students in the face of more powerful deterrents. This has been documented by Rosenblatt et al. (41) in a review of the effects that federal grants to schools had on the production of primary care physicians. Between 1977 and 1985, $461 million in federal grants was awarded to establish quality residency programs in family practice, general internal medicine, and general pediatrics, yet the proportion of graduates entering primary care remained stable over the next decade. Whereas it is true that those grants may have slowed the decrease in students choosing primary care fields, there is no evidence to document this without addressing the reason why students do not choose primary care fields. Restructuring internal medicine residencies or increasing family practice residents without diminishing the total number of available positions will not be effective.

Access to Medical Care

Finally, and perhaps most important, despite the adequate number of physicians in this country, including those in primary care, there are millions without access. However, even this issue becomes clouded by emotions rather than clarified by dispassionate discussion. It is alleged that more than 37 million people, through the lack of insurance or geographic location, have no access to medical care, with those in rural areas or in inner cities being particularly affected. In addition, there are felt to be millions more who, although covered by Medicaid, do not have access to primary care and are forced to go to emergency rooms to receive nonemergency treatment (42). This number, however, is truly an estimate and an inaccurate one at that (43). It includes those who, at one point in time, have no coverage but within a year may obtain insurance (estimated at 70% or 25 million people). It includes "repeaters," those who have insurance, lose it, and then regain it. It includes students with relatively minimal health care needs who will become employed and obtain insurance. In addition, there are many without insurance living in urban areas who can obtain care through municipal hospital systems that can provide fine, but not personalized, care, a goal to which some managed care groups may be striving in an attempt to be "cost effective." However, the actual number of those without insurance is not the point, for even if there were only 5 million people without insurance or access to health care, this is intolerable and must be addressed. This solution, however, should be one that maximizes, rather than diminishes, the chances for success.

The reasons why physicians do not wish to practice in underserved areas are quite clear, as are the realistic remedies to address not only this issue but the means of providing reimbursed care for those without insurance as well. It is clear that appropriate legislative remedies can be most effective. Such remedies should be targeted toward improving access, not increasing numbers of primary care physicians. As an example, in 1990, nine states (California, New York, Texas, Florida, Pennsylvania, Illinois, Ohio, Massachusetts, and New Jersey) were the sites for 54% of all actively practicing physicians (5). Yet, in each of these states considerable areas remain underserved. Ratios of physicians per 100,000 population vary from 181 in the East/South Central United States to 320 in New England. The number of physicians in practice, exclusive of residents, differs considerably in metropolitan statistical areas (MSA) as compared with non-MSA (309,566 vs 70,467). Yet, in every MSA pockets of underserved areas exist and, in most cases, are underserved by both generalists and specialists.

Clearly the best chance for success is to provide incentives to encourage physicians to practice in underserved areas. These incentives should be open to all rather than to specific groups. It should not be assumed that because one comes from a specific socioeconomic bracket or ethnic group, there is a special obligation to serve in a less desirable place or not to be a specialist. There should be more scholarship and loan-forgiveness programs open to those who not only wish to enter primary care but also agree to serve in underserved areas. To attract as heterogeneous a constituency as possible, such programs should not be need-based. The concept of the National Health Service Corps (NHSC) should be markedly expanded to allow underserved areas to receive an appropriate number of physicians. As tuition continues to increase, participation in the NHSC will become increasingly attractive. Petersdorf (44) has demonstrated that a person $75,000 in debt in 1990 would need to have an annual income of $146,000 5 years after graduating to repay this debt without severe financial distress. This exceeds the average income of a generalist who has recently completed his training. Although the commitment to the NHSC may only be 4 years, nonetheless, on a staggered basis quality care can be provided. In addition, if while serving in such a location an effort is made by the community to welcome the physician, he/she may well choose to stay after the term of service has expired.

Loan forgiveness, available upon graduating from medical school and based on career choice or ultimate practice location, can also be a powerful incentive. Recently, the New York State Department of Health announced the Loan Forgiveness Program that, depending upon site of residency training, could allow a resident to diminish indebtedness by $10,000 to $20,000 over a 3-year period. Programs such as these should be strengthened, with additional incentives provided for those willing to practice in an underserved area after finishing training.

It has also been suggested that universal service be required for all physicians, similar to the prevailing situation during the Vietnam War (44). This requirement would be accompanied by an essentially free undergraduate medical education. Although this concept does have appeal, as does universal service for all young adults, it is unlikely that "singling out physicians for service" could withstand legal challenge. A more reasonable, but expensive, proposal would be the expansion of available government funding to provide free tuition to all who agree both to enter a primary care field and to practice in an underserved area for at least 4 years. This, to some extent, is being accomplished through expansion and revitalization of the National Health Service Corp.

Conclusion

Our system of providing health care needs a critical review and considerable change. There is waste in the system, some charges for medical care are exorbitant, and too many people are without adequate care. We do have a physician surplus, and this must be addressed. However, this can best be accomplished and is in the process of being accomplished by downsizing the number of residency positions rather than by holding medical schools responsible for career choices. Our educational system for medical students needs revision, but primarily because of the changing patterns of medical practice.

With respect to physicians or those who wish to be, it has been my experience that virtually all students entering medicine are interested in helping people. Students are a heterogeneous group. Some are more technically oriented than others, some are more interested in lifestyle, and some are more committed to primary care. However, at present, the deterrents to those considering primary care, as well as to those who wish to practice in underserved areas, are considerable. As these deterrents are being removed, and the medical marketplace is making its preferences known, many more students are selecting primary care. If incentives are provided to practice in underserved areas, more physicians will serve.

We can learn from what has been done in other industrialized nations—but learn not only what to do but what not to do. It should be remembered that although 75% of the public is aware of a health care crisis, 85% are also satisfied with their own health care. Our goal must be to improve the care for those without, not lower the care for the majority of those whose needs are being met.

References

1. News from the Robert Wood Johnson Foundation. Tuesday, November 10, 1992.
2. Petersdorf RG. The doctor is in. Presented at the Plenary Session of the 103rd Annual Meeting of the Association of American Medical Colleges, New Orleans, Louisiana, November 6–12, 1992.
3. Report of the Graduate Medical Education National Advisory Committee, Volume II, U.S. Dept of Health and Human Services DHHS publication no. 1 (HRA) 81–652, 1980.
4. Sixth Report to the President and Congress on the Status of Health Personnel in the United States. U.S. Dept of Health and Human Services DHHS publication no. (HRS-P-OD-88-1), 1988.
5. Physician characteristics and distribution in the U.S. Chicago, IL: American Medical Association, 1992.
6. AAMC policy on the generalist physician. As adopted October 8, 1992. Acad Med 1993; 68:1–6.
7. The Statesman's Year-Book. Statistical and historical annual of the states of the world for the year 1992–1993. B. Hunter, ed. New York: St. Martin's Press, 1993.
8. World Almanac, 1993. New York: Newspaper Enterprise Assoc.
9. Stoddart GL, Barer ML. Toward integrated medical resource policies for Canada: 2. Promoting change—general themes. Can Med Assoc J 1992; 146:697–700.
10. Nash DB. White coat or blue collar. In: Nash DB, ed. Future practice alternatives in medicine, 2nd ed., New York: Igaku-Shoin, 1993: 1–29.
11. NRMP data. Evanston, IL: National Resident Matching Program, 1990, 1992, 1993.
12. AAMC Facts. Applicants, matriculants and graduates 1987–1993. Washington, DC: Association of American Medical Colleges, January 1993.
13. AAMC Reporter. Washington, DC: Association of American Medical Colleges, 1993; 2:7.
14. Koop CE. A general and primary solution for American health care problems. Acad Med 1993; 68:46–47.
15. Stimmel B. The crisis in primary care and the role of medical schools. Defining the issues. JAMA 1992; 268: 2060–2065.
16. Petersdorf RG. Financing medical education. Acad Med 1991; 66:61–65.
17. Levinsky NG. Recruiting for primary care. N Engl J Med 1993; 328:656–660.
18. Fuchs VR. No pain, no gain. Perspectives on cost containment. JAMA 1993; 269:631–633.
19. Supplying physicians for future needs. The Report of the Task Force on Physician Supply. Washington, DC: Association of American Medical Colleges, 1990.
20. Hughes JS. How well has Canada contained the costs of doctoring? JAMA 1991; 265:2347–2351.
21. Nutting PA, Franks P, Clancy CM. Referral and consultation in primary care: do we understand what we're doing? J Fam Pract 1992; 35:21–23.
22. Page L. Physician supply woes. Is Canada getting too ruthless with doctors? Am Med News, May 17, 1993, p. 1.
23. Seltzer I. There is no health care crisis. The Wall Street Journal, Tuesday, January 25, 1994, A12.
24. Manning WG, Leibowitz A, Goldberg GA et al. A controlled trial of the effect of a prepaid group practice on the use of services. N Engl J Med 1984; 310:1505–1510.
25. Egdahl RH, Taft CH. Financial incentives to physicians. N Engl J Med 1986; 315:59–61.
26. Himmelstein DU, Woolhandler S, and the Writing Committee of the Working Group on Program Design. A national health program for the United States. A physician's proposal. N Engl J Med 1989; 320:102–108.
27. Himmelstein DU, Woolhandler S. Cost without benefit: administrative waste in U.S. health care. N Engl J Med 1986; 314:441–445.
28. Woolhandler S, Himmelstein DU. The deteriorating administrative efficiency of the U.S. health care system. N Engl J Med 1991; 324:1253–1258.
29. Sack K. Audit faults market for ills of Blue Cross. New York Times, Sec 1:37, April 18, 1993.
30. Judis JB. Whose managed competition? New Republic, March 29, 1993, pp. 20–24.
31. Franks P, Clancy CM, Nutting PA. Gatekeeping revisited—protecting patients from overtreatment. N Engl J Med 1992; 327:424–429.
32. Murray D. Managed care: the payoff keeps getting better. Med Econ. December 7, 1992, pp. 124–142.
33. Calman NS, Hyman RR, Licht W. Variability in consultation rates and practitioner level of diagnostic certainty. J Fam Pract 1992; 35:31–38.
34. Davis K. Will managed competition work? Health/PAC Bull 1993; 23:25–26.
35. Rubin HR, Gandek B, Rogers WH et al. Patients' ratings of outpatient visits in different practice settings. Results from the medical outcome studies. JAMA 1993; 270:835–840.
36. Flexner A. Medical education in the United States and

Canada: A report to the Carnegie Foundation for the advancement of teaching. Washington, DC: Science and Health Publications, 1910; pp. 93–94.

37. Pawlson LG, Schroeder SA, Donaldson MS. Medical student instructional costs in a primary care clerkship. J Med Educ 1979; 54:551–555.
38. Stern RS, Jennings M, Delbanco TL et al. Graduate education in primary care: an economic analysis. N Engl J Med 1977; 297:638–643.
39. Kirz HL, Larsen C. Costs and benefits of medical student training to a health maintenance organization. JAMA 1986; 256:734–739.
40. Straub WH. Physician work force needs undermanaged competition: analysis and proposed management model. Wyoming: Jackson Hole Group, 1993.
41. Rosenblatt RA, Whitcomb ME, Cullen TJ et al. The effect of federal grants on medical schools' production of primary care physicians. Am J Public Health 1993; 83: 322–328.
42. Grumbach K, Keane D, Bindman A. Primary care and public emergency department overcrowding. Am J Public Health 1993; 83:372–377.
43. Swartz K. Dynamics of people without health insurance. Don't let the number fool you. JAMA 1994; 271:64–66.
44. Petersdorf RG. Financing medical education. A universal "Berry Plan" for medical students. N Engl J Med 1993; 328:651–654.

Editor's Comment

SHERMAN KUPFER, M.D.
Editor-in-Chief

Time-Warp Commentary; Back to the Future

"AN IMAGINARY CONVERSATION with the Gang of Three: A Ghostly Interview with Burrill B. Crohn, Leon Ginzburg, and Gordon Oppenheimer," an article by Dr. Janowitz that appears in this issue of the Journal, certainly is titled properly. I am sure that the readers will find the article interesting and entertaining, facets of medical journalism that are often in short supply. Dr. Janowitz knew Drs. Crohn, Ginzburg, and Oppenheimer very well and we are grateful indeed to have his creativity and imagination provide us with some novel insights into the background of the description of this important disease entity as he has so often done. I, too, knew Drs. Crohn, Ginzburg, and Oppenheimer and have had the pleasure and privilege of hearing them speak individually on the topic to a greater or lesser extent. Another reviewer for the Journal who had very close association with these individuals also felt that the article was worthy of publication in the Journal. Simply put, Dr. Janowitz has provided us once again with information that I would label as a time-warp commentary on the past and "going back to the future." Clearly, much more work needs to be done to unravel the underlying pathogenesis of this disease. I trust that this further effort is well served by this piece. Dr. Janowitz has told me that he intends to write a detailed documented history. Prior to that report, however, and in this respect, I think that some individuals may be interested in reading some of the articles for more than just historical purposes. To facilitate this interest, an abbreviated bibliography follows.

Bibliography

Crohn BB, Ginzburg L, Oppenheimer GD. Regional ileitis. JAMA 1932; 99:1323–1329.

Dalziel TK. Chronic interstitial enteritis. Br Med J 1913; 2:1068–1070.

Ginzburg L, Colp R, Sussman L. Ileostomy without exclusion. Ann Surg 1939; 110:648–658.

Ginzburg L, Oppenheimer GD. Non-specific granulomata of the intestines; inflammatory tumors and strictures of the bowel. Ann Surg 1933; 98:1046–1062.

Ginzburg L, Schneider KM, Dreizin DH, Levinson C. Carcinoma of the jejunum occurring in a case of regional enteritis. Surgery 1956; 39:347–351.

Hadfield G. The primary histological lesion of regional ileitis. Lancet 1939; 2:773–775.

Lee ECG, Papaioannou N. Minimal surgery for chronic obstruction in patients with extensive or universal Crohn's disease. Ann R Coll Surg Engl 1982; 64:229–233.

Lockhart-Mummery HE, Morson BC. Crohn's disease (regional enteritis) of the large intestine and its distinction from ulcerative colitis. Gut 1960; 1:87–105.

An Imaginary Conversation with the Gang of Three:

A Ghostly Interview With Burrill B. Crohn, Leon Ginzburg, and Gordon Oppenheimer

Henry D. Janowitz, M.D.

Through the good offices of Drs. Korelitz and Prantera, this unusual dialogue with these three physicians "on the other side" allowed me to discuss some aspects of "Regional Ileitis" with its original describers.

Henry D. Janowitz (HDJ): Good evening, Burrill, Leon, and Gordon. It has been a nearly a quarter of a century since I had the pleasure of seeing you all, in 1968, at a meeting of the National Foundation of Ileitis and Colitis of which you all were honorary chairmen.

Gordon Oppenheimer (GO): Someone joked that this was the first time we had been together again since we wrote the original paper.

Leon Ginzburg (LG): Like most jokes, it was probably true, but, like a lot of other things, we don't worry anymore about that over here.

HDJ: So I can call it Crohn's disease without stepping on anyone's toes?

Burrill B. Crohn (BBC): Now, Henry, you know I never referred to Crohn's disease. It was Brian Brooke in his editorials in the *Lancet* who popularized the term.

HDJ: Yes, Burrill. When I got to know you better in the 1950s, you always seemed almost embarrassed by the term. Yes, Leon, although you are smiling, that's true.

Now that I have got you all here together, there are several things I want to get straight. It was early in 1960 that Lockhart-Mummery and Morson put regional enteritis of the colon definitively into the literature and on the map.

Now Leon and Gordon, you had described the colonic involvement of granulomatous disease in your first independent paper in the surgical world in 1933 after the joint paper in New Orleans in 1932. Why didn't you follow up on that?

LG: The idea seemed clear enough to me and obvious. Gordon was interested in urology and I was born to be a surgeon. Pathologists and medical men could worry that bone.

HDJ: I can remember almost daily conversations with you, Burrill, Dick Marshak, our great radiologist, and me as a newcomer about the fact that the same process that occurred in the ileum could occur in the colon. I can recall how exasperated Dick was with you and your reluctance to accept the concept.

BBC: Yes, Dick Marshak was easily exasperated. But you remember I did come around to accept the fact although a bit slowly.

HDJ: I always blamed that on the great hold that the Department of Pathology under Paul Klemperer had on medical conceptual thinking at Mount Sinai.

BBC: You are right. Dr. Otani was also a great pathologist, but he clung firmly to the belief that the ileocecal valve separated ileitis from ulcerative colitis. Ileitis north, ulcerative colitis south of the valve. Others like Wells had been writing

---

Adapted in part from a chapter that will appear in a forthcoming book by C. Prantera and B. I. Korelitz, *Crohn's Disease*, New York: Marcel Dekker Inc., May 1996.

From the Dr. Henry D. Janowitz Division of Gastroenterology, The Mount Sinai School of Medicine, New York, New York.

Address reprint requests to Henry D. Janowitz, M.D., Mount Sinai School of Medicine, Division of Gastroenterology, Box 1069, New York, NY 10029.

for some years, long before Dick Marshak, and finally Lockhart-Mummery and Morson published their paper.

HDJ: Now something else has bothered me for years. You published your original paper from the Medical and Surgical Services and the Department of Pathology. Why didn't it have a single photograph of a gross specimen, or a single photomicrograph?

LG: My hazy recollection is that Gordon and I published quite a few gross anatomic pictures in our own paper and I suppose the Department of Pathology was planning to publish the histology for the disciples of that field.

HDJ: So the emphasis on the granuloma system became important only after Hadfield's paper 7 years later.

I know according to the rules I am supposed to ask the questions but surely there is one most all of you are burning to ask.

BBC, LG, GO (simultaneously): Of course! Have you found out either the cause or the cure by now?

HDJ: Unfortunately, we still do not know the cause or have a cure.

LG: I may be less cynical and also less sarcastic on this side, but I have not lost my wits. I always felt that if we didn't find out right away that it would take a long time.

HDJ: It has taken a long time, but perhaps 62 years is not really a long time. We have learned a lot about the natural history and the manifestations of the disease and its pathophysiology, we have accumulated a lot of information regarding the intimate steps of the inflammatory process of the bowel, quite a few medicines that help, and we have come to appreciate the gains and losses of surgery.

What I want to do now is compare what you thought then with what we have added or subtracted since your clinical description of this disorder. I always thought that there were two pleasures of growing old in medicine, I didn't say in growing old, but growing old in medicine. First, one learned what had happened to one's former patients. Second, the outmoded theories of our youth are revived.

In this context you will remember, Burrill, you told me that when you first met me in 1939, 7 years after the seminal paper, that Emanuel Libman, distinguished clinician whose name is attached to Libman-Sachs disease, verrucous endocarditis, had proposed rather facetiously that ileitis was a new disease due to the widespread use of toothpaste containing silica particle matter.

BBC: But I don't think that this was taken very seriously.

HDJ: No, but the Chesses did, and fed silica-containing mixtures to animals and produced lesions of the small bowel. However, you may be interested that during the last year or two a series of letters in the *Lancet* had revived the theory and paid attention to the possible submicroscopic material in the intestine lesions. But I too feel this will not lead us very far.

But in a more serious vein, what did you three or perhaps the original surgeon, A. A. Berg, think about this disease when you first began to study it and recognized its distinction from other small bowel conditions?

LG: At the operating table it looked different from our preoperative expectations of either cancer or tuberculosis and certainly when we opened the specimen. One couldn't help thinking that it might be a variant of tuberculosis in these relatively young people.

BBC: One couldn't help thinking of other granulomatous diseases of the area—tuberculosis, actinomycosis, or some chronic infection.

GO: It was obvious very soon that the lesions did not contain acid fast organisms, but we were all excited about an infectious etiology. So we injected lots of material from lots of specimens into lots of animals and never produced anything resembling regional ileitis.

HDJ: This idea of an infectious cause has never disappeared from the scene and has been pursued very vigorously and revived from time to time. The search in recent years has been for a mycobacterium, because the veterinary world had soon recognized the gross similarity of regional enteritis to Johne's disease of cattle caused by a paratuberculosis mycobacterium.

The literature on the possible mycobacterial origin of Crohn's disease is tremendous. Many clinical trials of treatment with combinations of up to four or five antituberculosis drugs have had varying results, but I think it is fair to say at present that the case is still not proven.

The workers are at it strong and heavy using more refined technology to demonstrate biochemical residuals of mycobacterial organisms in the tissues. You should know that the organism producing Whipple's disease seen by light microscopy and by staining techniques in 1902 has finally been identified by the newer techniques.

LG: We also heard that this was accomplished just as Whipple's disease began to disappear as an intestinal disorder.

HDJ: But as more workers devoted themselves to this search for the infectious agent and none was forthcoming, what did you think, Burrill?

BBC: I still thought it was caused by an infecting agent. Some patients got better on the sulfa-containing drug, sulfathaladine, which was used widely by the profession before the introduction of steroids, but we never had the use of antibiotics directed against intestinal organisms. Then the introduction of steroids for inflammatory bowel disease, which seemed to be so useful, turned our thinking in other directions.

HDJ: Even in the late 50s you kept thinking in terms of "a viral etiology."

BBC: Yes. I always preferred a simple to a complex answer to a problem. I thought it might turn out to be a virus.

HDJ: I clearly remember your responding to my question as to whether Crohn's disease and ulcerative colitis, which you clearly separated, could occur in the same individual.

BBC: Don't laugh. I said yes. If the agent is possibly a viral one and why couldn't a patient with one virus have been infected by another virus and have both diseases?

HDJ: Another early concept that you, Leon, stressed is having a renewal of life. I mean the idea that the fecal stream plays a role in the pathogenesis of Crohn's disease. You recall, Leon, that you came to this concept from your observations on these results of diversionary operations in regional enteritis.

LG: Thanks for reminding me. The observations were made so long ago that I have almost forgotten them. In the 30s operations for regional ileitis were not risk-free. We had no pre- or postoperative access to antibiotics, so removal of disease was done in two stages. In the first operation the ileum was divided proximal to the disease and then closed. The proximal end was then implanted in the colon: the usual ileocolostomy. Several months later in the second operation, the diseased ileum and cecal tip were removed. I was impressed how often the bypassed ileum had become a shrunken fibrotic cord and how much less inflamed. I had never seen this in the disconnected rectum of ulcerative colitis after subtotal colectomy and it impressed me.

HDJ: I never forgot you telling me this and I tried to explain the ileal and anal location in Crohn's disease as being related to the pile-up of the fecal stream at the valvular areas. So you may be interested in the fact that recent studies once again suggest a role for the fecal stream. One set of workers have temporarily protected an ileal ascending colostomy for regional enteritis by a proximal ileostomy for up to 6 months and have not seen the usual endoscopic inflammation that usually takes place. Another group has once again studied the reparative process in Crohn's disease of the colon after temporarily diverting the fecal stream with evidence of healing.

LG: Of course I am delighted to hear about this finding. The healing of granulomatous disease of the colon does not last very long. By 6 months patients were ready to have their colons removed.

BBC: But even in the days when we didn't separate the two causes of colonic inflammation occasionally an ileostomy led to temporary improvement.

HDJ: Didn't you sense that it was no accident that regional enteritis was frequently seen at a hospital like Mount Sinai, recorded by doctors with names like Ginzburg, Oppenheimer, and Crohn, and patients with Jewish names?

GO: At first we blamed it on the fact that we were a Jewish hospital, but later on we sensed that this was not the basis of this congregation of patients.

BBC: We knew right away that Bargen at the Mayo Clinic was seeing similar patients and collecting them.

LG: Probably patients from New York seeking a second opinion.

HDJ: By 1939–1941 during my internship we all recognized that it was a Jewish disease and joked about whether to blame the Jewish mother, Jewish food, or Jewish genes.

GO: It is our turn to ask questions. What did you learn about this congregation of patients?

HDJ: We are no longer the chosen people but many of us are called. Crohn's disease is a worldwide disease of all social classes and of all races. Much good data on the familial susceptibility to inflammatory bowel disease comes from cohort studies and twin studies. If you have Crohn's then there is a good chance that others in your family will too, and if you and your wife have the disease a large proportion of your children will too. But the problem still remains on how to identify the genetic markers. Some very exciting recent studies indicate that ulcerative colitis and Crohn's disease have different HLA typing, but not all fit into two distinct patterns. Equally exciting is the finding that patients with ulcerative colitis to a high degree but not all have a special marker in the blood known as ANCA (peripheral antinuclear cytoplasmic antibody), which patients with Crohn's disease do not. Don't ask me to define this further. I'll have trouble doing so.

BBC: I am delighted to hear how the search thrives. You will remember, Henry, that in 1961 you introduced me to a session of the New York Academy of Sciences devoted to granulomatous disease. The next day I had to confess to you that I really understood very little of what was presented.

HDJ: The important point of this result is that it opened the question of the unity of each disease. We spent the last 30 years separating nonspecific inflammatory bowel disease into two separate containers—ulcerative colitis and Crohn's disease. Now it would appear that there is heterogeneity within each group.

You all know that even with the specimen in their hands, expert pathologists have trouble classifying 10%, maybe 15%, "the indeterminate group." Some patients may lack both markers—HLA type and ANCA marker. So Burrill and Sidney Truelove might be right. Patients might have both diseases.

HDJ: I want to ask you one question from an evolutionary point of view. When you identified regional enteritis, did you think it was a new disease or one that had been overlooked in the past?

LG: I thought it was relatively new. We couldn't believe that the great pathologists from the nineteenth century, Virchow, Connheim, etc., who did routine examinations of the entire small bowel in every autopsy would have missed it. They knew tuberculosis very well and wouldn't have confused it.

HDJ: Even if it is only relatively new, not many observers have recognized it but one had. You will be interested in this, Leon. Dalziel, a prominent Scotch surgeon in 1912, published a paper in the *British Journal of Surgery* that clearly distinguished a group of patients with fibrosing lesions in the small and large bowel that have all earmarks of regional enteritis and Crohn's disease.

I asked you once before in another existence, Burrill, whether you and Leon had ever heard of this paper because Brooke specifically raised the question.

BBC: Paul Klemperer's Department of Pathology had long been interested in fibrotic lesions of the bowel. We had a very good library in those days but never came across Dalziel's paper.

HDJ: To me the most unusual question is why didn't the British surgeons pick up this report and pay more attention to it? It was not published in an obscure journal.

LG: Part of the answer, I believe, is that there were no further follow-ups to create the appropriate receptive climate like the original JAMA paper by us three and the surgical paper by Gordon and myself in 1932. The Mount Sinai group, especially Burrill, kept on collecting and writing about the entity. Except for the paper on cancer of the jejunum in regional enteritis, I don't think I wrote any more and I am pretty sure that Gordon didn't.

HDJ: Which brings me to this following question. How important did you think your original observations were or might turn out to be?

GO: I thought it was an interesting finding, but I went on to fry other fish.

LG: It was new, interesting, even curious. It might turn out to be important, but I didn't give it that much thought. Remember, in those days we were supposed to find new diseases at Mount Sinai; Koplik had his spots, Tay and Sachs had their disease, Brill described his disease, and there were a bunch of others.

BBC: It was new to us and exciting because of that. We didn't know whether it would be important. Others like Bargen were collecting their own cases. We thought we should keep on looking for new cases, keep our records separate, and I segregated the charts in my office. I didn't think the disease would take on such worldwide clinical importance. I was astounded after the flood gates opened with identification of involvement of the large bowel by Lockhart-Mummery and Morson.

LG: I was gratified by the confirmation of Gordon's and my observation that the colon was involved in this disease, but it was old hat by then to us.

HDJ: There was another observation that you, Leon, made fairly early that has blossomed forth as an important point for the future.

LG: I didn't write much about regional enteritis but you remind me that I did report the first case of carcinoma of the small bowel and the jejunum in a patient with long-standing ileojejunitis.

HDJ: Yes, 22 years after the original description of regional enteritis you presented a paper on a patient with regional enteritis who had a cancer of the small bowel in 1955. You obviously were aware of the propensity of patients with ulcerative colitis to develop cancer and you explained the rarity of your case on the basis of the fact that cancer rarely occurs in the small bowel. You were right about the rarity, but since then almost all the cancers of the small bowel have occurred in patients with Crohn's disease. For a while we thought they were on the bypass loop and that played a role, but it is clearer now that cancer takes a longer time to develop than it does in ulcerative colitis. It has little to do with the bypass.

LG: What is the story in Crohn's disease in the colon?

HDJ: It is far from settled but the likelihood of cancer in the colon is increased several-fold over the normal expected rate in the general population, but perhaps not as often as in ulcerative colitis in the colon.

Our current belief is that few patients tolerate regional enteritis of the colon without colectomy for long periods of time. Most observers, although not all, feel that cancer is more likely in Crohn's of the colon and takes a decade more to develop than in ulcerative colitis.

BBC: Do you keep all the patients with Crohn's of the colon under surveillance with colonoscopy at the present as you do with ulcerative colitis?

HDJ: The consensus is not in yet although many of us do.

LG: The one case I reported seems to have started up something.

HDJ: But not as much as the 14 original patients with regional enteritis. Let me ask all of you this. Did the problem of cancer in regional enteritis, small or large bowel, occupy much of your concern?

LG: We really didn't believe that we gave the cancer problem much thought. My case just seemed an oddball.

HDJ: As we approach the end of this interview, I want to direct comments to each of you that would indicate that we have made some progress in understanding the entity you so clearly described 63 years ago.

Gordon, the urologic aspects of this intestinal disease have wide spread ramifications. We understand the nature and therapy of uric acid calculi, oxalate metabolism, and the role of hyperoxyluria in relation to the formation of oxalate stones. Obstructive uropathy and bladder fistula have given an important place to the urologist in the management of enteritis of the large and small bowel.

GO: Had I but known . . . .

HDJ: Leon, we understand now why your ileitis patients had gallstones about which they kept telling us, secondary to the reduction of their bile salt pool. But saving bowel is the watchword of small bowel surgery today with the use of a stricturoplasty instead of resection whenever possible.

LG: I await the long-term follow-up.

HDJ: Burrill, as the colonic involvement of the basic disease became clearer, a whole host of extraintestinal manifestations became obvious. They included ileitis, stomatitis, axial and peripheral arthritis, erythema nodosum, pyoderma gangrenosum, even pancreatitis. The spectrum has widened and each manifestation became more puzzling, which has made the whole range of this disorder more interesting.

BBC: So Crohn's disease has replaced the Lues of my medical student days.

HDJ: Although the whole world has paid tribute to your joint papers, miracle of miracles, you have not been forgotten at your own hospital. Burrill, you have been honored by a medical professorship linked to gastroenterology and, Gordon and Leon, you are honored each year by an annual lectureship at your medical school established by your friends and colleagues, given by distinguished members of the profession.

And, of yes, the National Foundation of Ileitis and Colitis, now renamed the Crohn's Colitis Foundation of America, of which you were the honorary chairmen, thrives, having raised and spent about 27 million dollars of funds for teaching, education, and research. It serves as the model for similar foundations throughout the world.

I hope the next time we meet I can bring the news that we have found the cause, or at least a cure, for this disorder.

Book Reviews

How to Quit Smoking without Gaining Weight

Martin Katahn, Ph.D., New York: Norton & Co., 1994. 223 pp, $12.95. ISBN 0-393-03714-2.

This book addresses an important issue not usually covered in the many self-help books on smoking or nutrition—how to simultaneously modify two of the behaviors that are the most difficult to change: smoking addiction and diet. The premise of the book is sound. Numerous anecdotal and clinical studies document that weight gain accompanies smoking cessation; hence there is a need for practical advice for the millions of smokers who wish to give up their addiction but are afraid of gaining weight as a result of their efforts.

The method by which Dr. Katahn proposes weight reduction is called the 7 + 7 Program. This is a two-week program. The initial week focuses on self-observation and psychological preparation. The second week consists of a gradual reduction in the number of cigarettes smoked. A parallel diet and exercise plan is coupled with the cigarette reduction plan.

Much of the general advice given in this book is good common sense. The diet and exercise programs appear to be based on the author's expertise in nutrition and weight loss, but the sections dealing with pharmacotherapy are sketchy, and many supportive measures for smoking cessation are not discussed. The most troubling aspect of this book is a lack of follow-through for the smoking program. Most smokers can manage to reduce their dependence on cigarettes for short periods of time, but the relapse rate can be as high as 90%. Unfortunately, this book offers little that can alter this fact.

Although many useful tables and charts are included and some interesting cases are discussed at length, no clinical evidence that the 7 + 7 approach is effective is presented. This book may be helpful for weight control to the smoker who has already quit, but it does not offer very much for the current smoker.

E. Neil Schachter M.D.
Mount Sinai Medical Center

The PDR Family Guide to Women's Health and Prescription Drugs

Medical Economics Data Production Company, Montvale, NJ, 1994. 891 pp, $24.95. ISBN 1-56363-086-9.

Medical Economics, the publishers of *The Physician's Desk Reference,* have released another volume in The PDR Family Guide Series. The first volume was *The PDR Family Guide to Prescription Drugs.* Their latest release, *The PDR Family Guide to Women's Health and Prescription Drugs,* is a near 900-page comprehensive text. The first half deals with the unique health concerns of women. General health concerns which include heart disease, AIDS, headache, and cosmetic surgery are discussed from a gender-specific perspective. The text includes a brief section on nutrition and health. Information is presented on basic normal physiology and signs and symptoms of the normal and diseased states. The second half of the book provides complete prescribing information on a variety of medications commonly used by women.

The aim of the book is to provide education about the female genitourinary tract in health and disease. Virtually every symptom, disorder, or disease related to the reproductive tract is discussed in depth, including some basics about normal physiology. Specific reproductive issues including infertility, contraception, and pregnancy are discussed extensively. Expectations for menopause and changes associated with estrogen deficiency are dealt with adequately. Psychosocial issues and psychiatric disorders including stress, anxiety, depression, and eating disorders are discussed, as well as sexual assault. Other medical conditions that are specific to women are addressed, including breast, cervical, endometrial, and ovarian cancer.

The book is organized by symptoms or signs, and it includes normal anatomy and causes of various diseases. It goes into depth about methods to diagnose diseases by laboratory testing and explains the risks of the tests, how they are performed, the necessary preparation for them, and time needed to perform them. It even discusses the test interpretations, with false positives and negatives. It goes into graphic detail about various medical diseases and procedures, including pictures of the normal, dysplastic, and cancerous cervix and the instruments and procedures used to perform biopsies. Treatments are described including the reasons for treatment, length of treatment, and risks and side effects.

Good health maintenance and disease prevention measures are dealt with appropriately. Women are encouraged to help themselves with instruction on nutrition and exercise. The book is supportive of physicians and always stresses the need for doctors' advice, diagnostic skills, and treatment choices.

It is not clear who would benefit best from the book. The language is at times formal and too scientific. When discussing treatment regimens for hypertension, the terms *vasodilators, ACE inhibitors,* and *antagonists* are used without explanation. The depth of information in this book will be useful for health care professionals, but might overwhelm the general public. Upon reading this book, patients' minor symptoms and signs could be misinterpreted and raise fears of serious medical conditions. Instead of being reassuring, as suggested in the foreword, this book could alarm lay readers unnecessarily.

This book gives extensive information on a variety of gender-specific states, but other significant information for women is missing, such as substance abuse (cigarettes, alcohol, and drugs), domestic violence, and autoimmune disorders—problems equally germane to women. The other problem is misinformation. In discussing treatments for the premenstrual syndrome, L-tryptophan is mentioned. This drug was banned several years ago by the Food and Drug Administration because of deaths due to the eosinophilia-myalgia syndrome. In addition, when discussing osteoporosis, the book states that the best test to measure bone density is dual photon absorptiometry. For several years, the most accurate and precise method has been dual energy x-ray absorptiometry.

The second half of the book is an alphabetical listing of the drugs most commonly used by women. The descriptions include pronunciation, indications, method of intake, side effects, risks, and contraindications. Also included are food and drug interactions, information about use in pregnancy and breast-feeding, dosage, and overdosage. The drugs are cross-referenced by generic name. The information is useful, concise, and well written, without being alarming. Only significant side effects are included.

Overall, this is a useful text with a wealth of information about medical conditions related to the female genitourinary tract and some other female gender-specific illnesses. It also provides the reader with means of disease prevention and maintenance of good health. Although it is a bit dry and reads like a medical text in some sections, it is concise and well written and makes the reader an educated consumer of health care. This book is for the patient who needs to know and have control over every aspect of her medical care.

Erica Asch, M.D.
Mount Sinai Medical Center

Eccentric and Bizarre Behaviors

Louis R. Franzini, Ph.D. and John M. Grossberg, Ph.D., New York: John Wiley & Sons, 1995. 244 pp, $18.95. ISBN 0-471-54520-1.

ALTHOUGH THIS BOOK was written for the general public and eschews technical medical terms and avoids explanatory mechanisms as to the etiology of the bizarre and interesting behaviors it describes, it is an important treatise for the medical profession as well. Specifically, it exposes readers to ways of the mind that seem more farfetched than fantastic and to the wonders of an organ—the mind—that is influenced greatly by both its genetics and its environment. It is not a textbook, however.

The book contains 12 chapters—really 12 bizarre behaviors—that are not only easy to read, but graphic in their presentation: Erotomania; Münchhausen's Syndrome by Proxy; Frottage (pleasure through rubbing any of another's body parts); Vampirism; Sex Slavery; Autoerotic Asphyxia; Capgras and Other Misidentification Delusions; Folie a Deux; Koro; Savants; Trichotillomania; and last but not least, Necrophilia. Clinicians can identify these patterns of behavior according to the descriptions when they encounter these patients.

But then comes the rub: Some of these behaviors reach the level of a diagnosis, for example, trichotillomania, Münchhausen's syndrome by proxy. (These disorders even have psychiatric Diagnostic and Statistical Manual IV diagnostic code numbers). Some of the behaviors described are symptoms of other disorders. These two approaches mix apples and oranges. Erotomania (obsessive loving) can be a manifestation of a paranoid delusion, a manic-depressive disorder, an obsessive disorder delusion, etc. That is, there is a mixing of behaviors, thought patterns, with diagnoses. Notwithstanding, the authors' mode of presentation is "catchy" and arouses our innate voyeurism, especially for the bizarre and unthinkable. However, switching hierarchical levels—symptoms versus clinical syndromes—inflicts critical dilemmas for the clinician, and, in particular, for the individual who is not in the mental health domain and who is less likely to denote that a behavior is a symptom of another diagnosis or of many diagnoses.

In the medical model—and psychiatric disorders lie in that domain—it is essential to establish a diagnosis before treatment is undertaken. Diagnosis should guide treatment! This book does not guide the clinician to the underlying diagnosis(ses), but leaves the patient description at the symptom level. This does not optimize treatment decisions, which are enhanced by translating behavior(s) to the level of a disorder when possible. For example, with erotomania, is the clinician dealing with a true paranoid disorder, a schizophrenic disorder, or an obsessive-compulsive disorder? Although the thoughts and behavior may look similar on the surface—the manifest level—they may have disparate psychiatric etiologies—the latent level—mandating different therapeutic approaches. The entities are not ends in themselves, as the authors would lead the reader to believe.

A further difficulty is that certain behaviors (e.g., in Münchhausen's syndrome by proxy, the proposing that someone else has symptoms and an illness, for example, a child or an elderly parent) does not assist the clinician in differentiating this behavior from malingering, in which the driving force is conscious secondary gain. In Münchhausen's disorders perpetrators of the behaviors are conscious of what they are doing, but are often unaware of the reasons or motivation for their behavior.

This reviewer would have appreciated the authors referring more to past writers of the sensational and bizarre to put their book into an historical perspective: Kraft Ebing, *Psychopathia Sexualis* (1); the monograph *Some Uncommon Psychiatric Syndromes* by Enoch, Trethowan, and Barker (2); and Freud's description of the underpinnings of bizarre behaviors.

It is a tribute to *The Mount Sinai Journal of Medicine* that it sees the importance of informing the medical community of behaviors that may confront the clinician and that need further exploration and understanding. The bizarre becomes meaningful when its etiology can be dissected from the symptoms and behaviors that arouse our interest. Meaningful interventions that may curb destructive and, at times, life-threatening behaviors should be developed at greater length. The authors have fulfilled their purpose in bringing the eccentric and bizarre human behaviors to our attention in such a readable and, at times, entertaining volume.

References

1. Kraft Ebing. Psychopathia sexualis.
2. Enoch MD, Trethowan WH, Barker JC. Some uncommon psychiatric syndromes. Bristol: John Wright and Sons, Ltd., 1967 (The Stonebridge Press).

James J. Strain M.D.
Mount Sinai Medical Center

The Living Heart Brand Name Shopper's Guide

Michael E. DeBakey et al., New York: Mastermedia Limited, 1993. 448 pp, $14.95. ISBN 0-942361-74-1.

The Living Heart Brand Name Shopper's Guide serves as a reference to consumers who explore the supermarket in search of healthy foods. It is not for individuals who dine out extensively; for this, the authors refer the reader to their other book, *The Living Heart Guide to Dining Out.*

The first 90 pages are written in a textbook format with eight chapters of abridged descriptions about diet and heart disease, low-cholesterol eating patterns, dining out, weight control, high blood pressure, diabetes, dietary supplements, and label reading before explaining how to use the book. Although the chapters are well written, much of the contents are oversimplified, often controversial, and confusing. For example, the only mention of alcohol defines serving sizes of wine, beer, and hard liquor. The authors could have discussed the relationship between alcohol consumption and high-density lipoprotein cholesterol levels to meet the needs of informed readers.

The chapter entitled "Planning a Cholesterol Lowering Eating Pattern" describes two methods of controlling fat intake. The first, the servings of food method, is almost identical to the American Diabetic Association's Exchange Lists, which requires nutritional counseling for interpretation and is often difficult to apply. The second, the grams of fat method, uses fat gram counting, which is appropriate for self-instruction.

The book's core lists approved brand name and generic products divided into 22 sections with calories, fat, unsaturated fat, cholesterol, and sodium values. The letter "d" is placed next to foods suitable for diabetics. Each section begins with a table of food high in fats labeled "not recommended for frequent use." This is helpful for those who manage their diet by the avoidance technique.

The book's purpose to assist shoppers by identifying foods low in saturated fats is well served after page 88. The book is a responsible effort by a team of professionals. Their pledge to update annually due to the volume of new and revised low-fat foods strengthens their dedication. I would recommend this book to patients or physicians who want to reduce the time spent in their supermarket without compromising healthy choices.

Lori Silver, R.D., M.P.H., C.D.E.
Cardiovascular Institute
Mount Sinai Medical Center

Analyzing Survival Data from Clinical Trials and Observational Studies

Ettore Marubini and Maria Grazia Valscecchi, Chichester, England: John Wiley & Sons, 1995. 414 pp, $63.95. ISBN 0-471-93987-0.

This book is intended for researchers involved in the analysis of survival data collected in clinical trials and observational studies. It aims to be a practical guide for clinical investigators who have limited statistical knowledge, but it also provides a fairly rigorous outline of the underlying statistical principles. The book grew from previous lectures to physicians with some experience in statistics, and biostatisticians without experience in survival analysis.

The book succeeds for both audiences, although the time commitment to master the material for either will be substantial, as is unavoidable if one wants to go from practically no knowledge in the field to the boundaries of current applied research. The cornerstone of the text is, as is appropriate for a clinical audience, nonparametric survival procedures, such as the Cox hazards model, as opposed to parametric models. If one wants simply to understand the most basic computer output from such analyses, the investment of time to read the book may be difficult to justify. If one wants to prod one's statistical colleagues to do more than give basic analyses, or perhaps perform more sophisticated analyses including checking assumptions, using diagnostics, and constructing alternative models, the substantial time investment is justified.

The book is much more accessible to the lay reader (or even many applied biostatisticians) than other books designed for the statistical reader. However, the main advantage of the book to clinical researchers is the detail given and the lack of previous knowledge assumed in two areas: clinical trials and regression analysis. In terms of clinical trials, the beginning of the book gives a detailed introduction including randomization, the intent to treat principle and interim analyses, whereas the last chapter discusses meta-analysis. The main body of the book gives detailed analysis of several data sets with emphasis on implications of results. In terms of regression, basic concepts of regression such as interaction are discussed in detail rather than assumed known, as might be the case for other statistical textbooks on the subject.

A modest defect is that the one word this reviewer tried to find in the index, "interaction," did not appear, despite the appearance of this term in the book in several contexts. A cursory examination of the rest of the index indicated fairly broad coverage, so the 0/1 success rate may be a fluke.

Sylvan Wallenstein Ph.D.
Biomathematical Sciences
Mount Sinai School of Medicine

Instructions for Authors

THE MOUNT SINAI JOURNAL OF MEDICINE, established in 1934 as *The Journal of the Mount Sinai Hospital,* is a peer-reviewed general medical journal. It is indexed or abstracted in *Index Medicus, Current Contents, Science Citation Index, Hospital Literature Index, International Nursing Index, Excerpta Medica, Chemical Abstracts, Biological Abstracts,* and major databases. Theoretical and basic science articles for the clinician, reviews, reports of clinical or research experience, articles on any subject affecting the practice of medicine, short notes in any specialty, book reviews, letters, and articles by medical students are welcome. The *Journal* is a participant in the Agreement on Uniform Requirements for Manuscripts Submitted to Biomedical Journals. The *Journal* aims at responding to authors of manuscripts within 8 weeks of date of submission.

Communications. Address correspondence on manuscripts review to: Managing Editor, *The Mount Sinai Journal of Medicine,* Box 1094, 50 E. 98th Street, 1E, New York, NY 10029-6574. Phone: (212) 241-6108; FAX: (212) 722-6386.

Authorship Responsibility, Financial Disclosure, and Assignment of Copyright. Each author is required to sign the three statements on the accompanying page.

Manuscript Preparation. Submit manuscripts in triplicate, including three sets of illustrations. Manuscripts must be typewritten, **double-spaced throughout (including references, legends, and tables),** on one side of 8½ × 11-in. white bond paper, with generous margins. Number pages consecutively. Begin a new page for title page, abstract, first page of text, references, each table, and legends. **Computer-generated typescripts** must be double-spaced and must be easy for editors and typesetters to read. **Computer tapes or disks:** consult managing editor.

Title Page. On the title page, on separate lines (double-spaced), include (1) title of article, phrased as concisely as possible; (2) name of each author, including first name and degree; (3) name and address of departments and institutions from which the work originated; (4) acknowledgments of sources of support; (5) name, place, and date of any conferences at which paper has been delivered; (6) the statement **Address reprint requests to . . . ,** giving full name and address, with zip code, of the appropriate author.

Abstract. All articles should include an abstract of not more than 250 words. Abstracts of clinical and nonclinical studies should be divided into four paragraphs: Background, on the problem addressed in the study; Methods, on how the study was performed; Results, giving the salient results only; and Conclusions. Case reports and review articles should also include abstracts, structured in labeled paragraphs as authors deem relevant to the manuscript. The abstract should be double-spaced and be on a separate page.

Key Words. Supply 3 to 10 key words or phrases at the bottom of the title page, for use in indexing the article. Use Medical Subject Headings from *Index Medicus.*

Writing Style and Abbreviations. Authors should aim for conciseness and clarity without repetition. Avoid medical jargon, abbreviated phrasing, and obscure or newly coined abbreviations. Define all abbreviations on first use, except those for standard, universally recognized units of measurement and statistical terms. Spell out such terms as white blood cell, hematocrit, superior vena cava.

Units of Measurement. Use SI units, giving range or interval of normal values.

Editing. All papers will be edited to enhance conciseness and clarity without altering meaning and to insure conformity with journal style. The author is responsible for all statements in the article.

Illustrations. Photographs, line drawings, graphs, and charts should increase understanding of the text. Line drawings, graphs, and charts should be professionally prepared. All illustrations should be submitted in triplicate as sharp, high-contrast glossy prints. Photomicrographs must be 5⅛ in. wide. On the back of each print affix a gum label with the number of the figure (numbered consecutively in arabic numerals as cited in text), title of manuscript, name of senior author, and the word "top" with an arrow indicating top edge. Illustrations which are to appear together should be mounted accordingly as plates and should correspond to each other in size. Protect the prints by enclosing them in heavy cardboard; do not use paper clips. Photographs of patients can be published only if a copy of the permission is submitted with the photograph. Eliminate names of patients and hospital numbers from x-rays.

Legends. Legends should not duplicate the text. Double space. Number each legend to match the illustration it accompanies. Illustrations mounted as plates to appear together should be accompanied by a single legend identifying the separate elements by position ("upper left," and so on) or by letters. If letters are used in the legend, a corresponding letter must appear on the print itself. Identify all abbreviations. Indicate magnification and stain for photomicrographs.

Tables. Each table should be typed on a separate page, double spaced. Number tables consecutively as they appear in text, using arabic numerals ("Table 1"). Give each table a brief title that is self-explanatory without reference to the text and type it on a separate line. If abbreviations or symbols are used in the table, identify them in a footnote.

References. When citing references in your text, the first work cited is numbered 1, and succeeding references are numbered in sequence; enclose the citation number in parentheses and place it before any punctuation: **Cytomegalovirus (1), steroids (2), and recreational drugs (3) have all been implicated (4).** The reference list includes only works cited in the text, numbered in the order in which they are cited. Type, double spaced, following the general arrangement and punctuation style in the examples below (journal title abbreviations conform to *Index Medicus* style):

1. Brown LK. Sleep apnea syndromes: overview and diagnostic approach. Mt Sinai J Med 1994;61:99–112.

2. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. N Engl J Med 1991;324:424–428.

When a manuscript is accepted for publication, you will be asked to confirm the accuracy of all references.

Proofs. Proofs are sent to the corresponding author from the printer. Stylistic changes which do not alter meaning should not be made at this stage. Because of increased printing charges, the cost of excessive author's alterations beyond routine correction of typographical errors and major errors in data may be billed to the author. Proofs should be corrected and returned to the editorial office within 48 hours.

Reprint Orders. A reprint order form is sent to the corresponding author with proofs; return it to the editorial office.

SOUTHAMPTON
HOSPITAL

and

IMUGEN

Invite You to Attend a Conference on

"LYME DISEASE: New Knowledge About a No-Longer-New Disease"

Saturday, April 20, 1996 • 9:00 a.m. - Noon
Southampton Hospital
Parrish Memorial Hall
Corner of Lewis Street and Herrick Road
Southampton, Long Island, NY

3 Hours C.M.E. Credit from Suffolk County Medical Society

This program includes
two renowned experts in the field.

PROGRAM

ALLEN C. STEERE, M.D., medical researcher and physician, is internationally recognized for his studies of Lyme disease. It was while working for the Centers for Disease Control in Atlanta that Dr. Steere and his colleagues at Yale concluded that a number of children in Lyme, Connecticut who were thought to be suffering from juvenile rheumatoid arthritis, were affected by a new illness that they called Lyme disease. He continues to study arthritic and neurologic complications of Lyme disease and to evaluate treatments and is currently a professor of medicine and chief of rheumatology/immunology at the New England Medical Center and Tufts University School of Medicine in Boston, Massachusetts.

ANDREW SPIELMAN, S.B., S.D., M.A. (Hon.), is a pioneer in the development of methods of diagnosing vector-borne disease and for protecting people against infection. His recent efforts have largely been directed toward an analysis of the environmental factors that promote the emergence of Lyme disease and human babesiosis. He is the author of more than 250 publications and holds three patents. Currently, he is a professor of tropical public health at Harvard School of Public Health.

A panel interaction and an audience question-and-answer session will follow the presentation.

There is no fee to attend; however,
seating is limited so please reserve early.

For more information and to register, contact the
Office of Medical Affairs, Southampton Hospital,
at (516) 726-8305

LEVY LIBRARY, MT. SINAI MEDICAL CENTER
3 4805 0408354 5

THE
MOUNT SINAI JOURNAL OF MEDICINE

Forthcoming:

Grand Rounds

Targeted Gene Disruption with Antisense DNA: Research and Therapeutic Applications
Alan M. Gewirtz
University of Pennsylvania School of Medicine

The Supraventricular Tachycardias
M. Noelle S. Langan
The Mount Sinai Medical Center

Irritable Bowel Syndrome
Lawrence S. Friedman
Harvard University

The Heart and Thyroid Disease
Paul W. Ladenson
Johns Hopkins University

Theme Issues

Computers in the Doctor's Scientific Office
Editor: Ray Cornbill

Issues in Medical Ethics
1. Violence and Medicine
Editors: Daniel A. Moros, Rosamond Rhodes, and Allyson Robichaud

Cross-Cultural Issues in Psychopharmacology
Editor: John Herrera

Emergency Medicine
Andrew Jagoda
Lynne Richardson

Available now:

Maternal and Infant Health: New Perspectives
Kenneth V. Lieberman and
Charles J. Lockwood, editors
volume 61, no. 5, October 1994
88 pages **$15**

Update on Sleep Disorders
Lee K. Brown, editor
volume 61, no. 2, March 1994
96 pages **$15**

Environmental and Occupational Medicine in Clinical Practice
Richard P. Wedeen, editor
volume 62, no. 5, October 1995
96 pages **$15**

THE
MOUNT SINAI JOURNAL
OF MEDICINE

OLUME 63 NUMBER 2 MARCH 1996

MAY 28 1996

Issues in Medical Ethics

Violence and Medicine

Daniel A. Moros, Rosamond Rhodes, and Allyson Robichaud, guest editors

From the Mount Sinai School of Medicine (CUNY)

STATEN ISLAND
UNIVERSITY
HOSPITAL
Affiliated with The Mount Sinai Hospital
and Mount Sinai School of Medicine

Veterans
Affairs

THE MOUNT SINAI JOURNAL OF MEDICINE

The Mount Sinai Journal of Medicine is published by The Mount Sinai Medical Center of New York and has the following affiliates: Bronx Veterans Affairs Medical Center, New York; Elmhurst Hospital Center, Queens, NY; Queens Hospital Center, Queens, NY; and Staten Island University Hospital, Staten Island, NY.

Editor-in-Chief

Sherman Kupfer, M.D.

Editor Emeritus

Lester R. Tuchman, M.D.

Associate Editors

Harriet S. Gilbert, M.D.
Julius Wolf, M.D.

Administrative Assistant

Joseline Ilagan

Assistant Editor for Housestaff

Matthew J. Kapklein, M.P.H., M.D.

Assistant Editors

David H. Bechhofer, Ph.D.
Constanin A. Bona, M.D., Ph.D.
Edward J. Bottone, Ph.D.
Lewis Burrows, M.D.
Christopher P. Cardozo, M.D.
Mark R. Chassin, M.D., MPH, MPP
Joseph S. Eisenman, Ph.D.
Richard A. Frieden, M.D.
Frederick Friedman, M.D.
Steven Fruchtman, M.D.
Paul L. Gilbert, M.D.
James H. Godbold, Ph.D.
Joseph J. Guarneri, Ph.D.
Richard S. Haber, M.D.
Noam Harpaz, M.D.
Tomas Heimann, M.D.
Sheldon Jacobson, M.D., FACEP
Barry W. Jaffin, M.D.
Andrew S. Kaplan, D.D.S.
Samuel Kenan, M.D.
Suzanne Carter Kramer, M.Sc.
Mark G. Lebwohl, M.D.
Zvi Lefkowitz, M.D.
Kenneth V. Lieberman, M.D.
Lynda R. Mandell, M.D., Ph.D.
Steven Markowitz, M.D.
Bernard Mehl, D.P.S.
Leonard E. Meiselas, M.D.
Myron Miller, M.D.
Edward Raab, M.D.
Elliot J. Rayfield, M.D.
Allan Reed, M.D.
Clive Rosendorff, M.D., Ph.D.
Fred Rosner, M.D.
Allan E. Rubenstein, M.D.
David B. Sachar, M.D.
Henry Sacks, M.D.
Ira Sanders, M.D.
Martin H. Savitz, M.D.
Clyde B. Schechter, M.D.
Stuart C. Sealfon, M.D.
Michael Serby, M.D.
Phyllis Shaw, Ph.D.
Alex Stagnaro-Green, M.D.
Barry D. Stimmel, M.D.
Nelson Stone, M.D.
Max Sung, M.D.
Rein Tideiksaar, Ph.D.
Richard P. Wedeen, M.D.

Editorial Board

Barry S. Coller, M.D.
Barry Freedman, M.B.A.
Richard Gorlin, M.D.
Nathan Kase, M.D.
Panayotis G. Katsoyannis, Ph.D.
Charles K. McSherry, M.D.
Jack G. Rabinowitz, M.D.
John W. Rowe, M.D.
Alan L. Schiller, M.D.
Alan L. Silver, M.D.
Alvin S. Teirstein, M.D.
Rosalyn S. Yalow, Ph.D.

The Mount Sinai Journal of Medicine (ISSN No. 0027-2507; USPS 284-860) is published 6 times a year in January, March, May, September, October, and November in one indexed volume by The Mount Sinai Medical Center, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Subscription price: individuals (U.S. and Canada), $75 per year; libraries and institutions (U.S. and Canada), $115 per year; individuals (outside U.S. and Canada), $100 per year; libraries and institutions (outside U.S. and Canada), $140 per year. Single copies $20. New subscriptions begin with the first issue of the calendar year. Send notice of change of address 60 days before its effective date. Second-class postage paid at New York, NY and at additional mailing offices. POSTMASTER: Send address changes to The Mount Sinai Journal of Medicine, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Telephone number (212) 241-6108; Fax (212) 722-6386. Copyright 1996 The Mount Sinai Hospital.

DISCLAIMER: The statement and opinions expressed in the *Journal* by authors and contributors are solely their own and not those of The Mount Sinai Medical Center or The Mount Sinai Journal of Medicine. All advertisements are expected to conform to ethical standards, and their appearance does not imply endorsement or approval by The Mount Sinai Medical Center and/or The Mount Sinai Journal of Medicine of the products with respect to effectiveness, quality, or safety. The Mount Sinai Medical Center and The Mount Sinai Journal of Medicine do not hold themselves responsible for statements made by the authors or contributors, and for any injury to persons or property which may have resulted from ideas, products, and services referred to in articles or advertisements.

COPYRIGHT: Material printed in the Journal is protected by copyright. In accord with U.S. Copyright law, no part of the Journal may be reproduced, displayed, or transmitted in any form, electronic or mechanical, including photocopying or by any information storage or retrieval system, without prior written permission to do so. Permission should be requested from The Editor. The Mount Sinai Journal of Medicine, Mount Sinai School of Medicine, Box 1094, One East 100th Street, New York, NY 10029.

THE
MOUNT SINAI JOURNAL
OF MEDICINE

Volume 63
Number 2
March 1996

CONTENTS

PART I

Introduction *Daniel A. Moros, Rosamond Rhodes, Allyson Robichaud* 71

The Nature and Limits of Violence *Bernard Gert* 72

The Legal Obligation to Prevent Harm: The Tarasoff Rule and Its Limits *Daniel Givelber* 77

Autonomy and Violence *Bat-Ami Bar On* 82

Cross-Cultural Perspectives on Violence: Is Violence Always Violence? *Arnold R. Eiser* 87

The Psychology of Violence *Robert Jay Lifton* 90

Violence against Elderly People *Barbara E. Cammer Paris* 97

Violence against Infants *Katherine Teets Grimm* 101

Violence and Women: Case Presentation and Discussion *Stephanie V. Seremetis* 103

Violence in the Emergency Department *Sheldon Jacobson* 105

Violence and the Physician's Duty: Panel Discussion *Daniel A. Moros, Chairperson* 107

Epilogue *Rosamond Rhodes* 111

PART II

Mount Sinai Medical School Commencement Address—May 19, 1995 *Antonia C. Novello* 112

GRAND ROUNDS

The Heart and Thyroid Disease *Paul W. Ladenson* 118

Irritable Bowel Syndrome *Lawrence S. Friedman and Richard B. Lynn* 126

DERMATOLOGY NOTES

Vitiligo *Marsha Gordon and Mark G. Lebwohl* 134

OPHTHALMOLOGY NOTES

Myasthenia Gravis *Arif Khan* 135

CASE REPORT

Gastrointestinal Histoplasmosis in a Patient with Acquired Immune Deficiency Syndrome *Jamal Raza, Michael T. Harris, and Joel J. Bauer* 136

LETTER TO THE EDITOR

Prostate Cancer Screening *Todd Simon; Reply by Norvin Perez* 141

Introduction

Daniel A. Moros, M.D.
Associate Professor of Neurology
Rosamond Rhodes, Ph.D.
Director, Bioethics Education
Associate Professor of Medical Education
Allyson Robichaud
Medical Ethics Fellow
Mount Sinai School of Medicine

Physicians and philosophers have long been cognizant of the pervasiveness of violence in everyday life. The Hippocratic Oath, written at a time when medicine and philosophy were less disparate activities, explicitly focuses on the interface between the medical professional and various forms of violence and specifies limits on a physician's behavior. In the present era, we are involved in increasing public debate on both the definition of violence and the proper limits of violence. We argue about the propriety of killing animals for food or our treatment of them in medical research, about contact sports (especially boxing) and corporal punishment of children, about the portrayal and exploitation of violence in the media. Today the response of physicians and other health care professionals to violence is no longer treated as a private matter to be addressed by a voluntary oath. Society now calls upon physicians to report not only gunshot wounds and dog bites, but child abuse and elder abuse. Indeed, the reporting of spouse abuse may not be too far in the future.

Although egregiously violent behavior may be easy to identify once we are alerted to it, the boundaries of unacceptable violent behavior are often ambiguous. The disciplining of a child by a parent, the demands of a ballet teacher or a high school athletic coach, and the battling between an aged invalid parent and a frustrated devoted child may be judged remarkably differently by different health care professionals. Furthermore, in our multicultural society, there are times when a health care professional may be uncertain about the acceptable range of behavior for individuals coming from very different and unfamiliar backgrounds. Such ambiguous cases are encountered daily in emergency rooms, on the wards, and in private offices. Only carefully developed insights into the nature and frequency of violence and the acceptability of violence can helpfully inform the moral judgments that are increasingly viewed as a component of a health care worker's responsibilities.

The current issue of the Journal presents papers from our 1995 conference, Violence and Medicine. This was the tenth Annual New York Regional Conference on Medical Ethics sponsored by the department of Medical Education of the Mount Sinai School of Medicine. These conferences offer a forum for a multidisciplinary panel of physicians and other health care professionals, philosophers, lawyers, regulators, and legislators to explore an issue of current debate and social concern. Our goal is to stimulate scholarly discussion that can have a direct impact on the concerns of the practicing health care professional.

Issues in Medical Ethics

The Nature and Limits of Violence

BERNARD GERT, PH.D.

Abstract

I take advantage of a point made by H. J. Gert, in "Rights and Rights Violators: A New Approach to the Nature of Rights" (*The Journal of Philosophy* 1990;87(12):688–694) to provide the following definition of violence. "Violence is an intentional or knowing attempt by a moral agent to directly cause harm, i.e., death, pain, disability, loss of freedom or pleasure, to someone who is protected by morality without the consent of that person." I show how this definition applies to violence in hospitals and discuss some of the ethical problems created by violent patients.

VIOLENCE CAN HAVE TWO QUITE DIFFERENT MEANINGS, and, although it is important to keep these meanings distinct, they often go together. The primary sense of violence, or at least the one on which I am going to be concentrating, is an act of violence (assault) committed by one person upon another. The first sense of violence in the Oxford English Dictionary is "The exercise of physical force so as to inflict injury on, or cause damage to, persons or property; action or conduct characterized by this; treatment or usage tending to cause bodily injury or forcibly interfering with personal freedom." This is the sense of violence we are using when we talk of one person committing an act of violence against another. Violence in this sense always involves another person, someone whom we regard as being protected by morality; even for the destruction of property to count as violence it must be the property of someone else.

The secondary sense of violence need have nothing to do with persons at all, let alone the action of one person against another. The Oxford English Dictionary characterizes this as "Force or strength of physical action or natural agents; forcible, powerful, or violent action or motion." This is the sense of violence we are using when we talk about a violent storm, or of someone exercising in a violent manner. Violence in the first sense need not be violent in the second. A person can very calmly, and almost unnoticeably, slip some poison into a person's drink to kill him. This is certainly an act of violence in the first sense, but it is certainly not a violent act in the secondary sense. However, it is not uncommon for acts of violence to be violent, for someone to do violence to someone by acting in a violent manner.

Indeed, if a patient is acting in a violent manner, swinging his arms around violently, we may seek to restrain him to prevent him from committing an act of violence. Note, however, that in restraining a violent patient, in either or both of the senses of violence, perhaps by giving him tranquilizing medication, we may be committing an act of violence against him, although if we do it cleverly this act of violence need not be done violently. Sometimes, however, an act of violence must be done violently to prevent a violent patient from committing acts of violence. Violence is not always unjustified; indeed, health care workers may sometimes be morally required to use violence, although usually the less violence, the better.

I shall now concentrate on the first of the two senses of violence, doing violence to someone, and try to describe with greater precision what counts as an act of violence. Although I started by quoting from the Oxford English Dictionary, I do not believe that even the Oxford English Dictionary provides a precise account of the meaning of a word, e.g., it defines the word "dog" as "a quad-

---

Eunice and Julian Cohen Professor for the Study of Ethics and Human Values, Dartmouth College.

ruped of the genus Canis, of which wild species or forms are found in various parts of the world . . ." If we take this definition literally, it means that there cannot be a dog with three legs. Thus, I am not troubled by the fact that my account of violence is not identical with what I quoted earlier, although I hope you will agree that it is somewhat more precise. In any event, I am not concerned primarily with providing an account of an English word, but rather with the concept of violence, one which is common in all languages. Violence involves causing or trying to cause harm to someone, so it is important to clarify what counts as harm (1).

Harms can be classified in several different ways. (I do not claim that the way I classify the different kinds of harm is the only way.) I have found this classification quite useful, and it seems to be easily understood (2). In this classification there are five main kinds of harm, they are death (3), pain, disability, loss of freedom, and loss of pleasure. Each of these harms, with the possible exception of death, comprise several categories. Further, with the exception of death, each of these harms has degrees; there are more and less severe pains, more and less serious disabilities, greater and lesser losses of freedom and pleasure. I mention this at the beginning so that no one will think that, by mentioning the harms in a certain order, they are being ranked.

Harms all have something in common; all persons in so far as they are acting rationally avoid them, unless they have an adequate reason not to do so. They also avoid any significant increase in their risk of harm. This distinguishes harms from everything else; if something is not a harm a rational person need not avoid it, even without a reason, e.g., listening to a philosophy talk. It is, of course, perfectly rational to avoid listening to a philosophy talk, but it is also rational not to avoid it. On the other hand, it is always irrational not to avoid a harm, unless one has an adequate reason. A desire for a harm does not even count as a reason, so it cannot be an adequate reason; simply wanting to die, suffer pain, be disabled or lose freedom or pleasure does not make it rational to seek them without a reason (4).

The primary beliefs that count as reasons for not avoiding these harms are that someone, possibly oneself, will avoid one of these harms. The other beliefs that count as reasons are beliefs that someone, possibly oneself, will benefit, i.e., gain abilities, freedom, or pleasure. All other reasons involve these basic reasons (5). A reason counts as adequate when any significant group of otherwise rational persons consider the harm avoided or benefit gained to be at least as important as the harm suffered. This means that there is not a uniquely "correct" way to rank the five kinds of harms. Although all rational persons agree on what is a harm and what is a benefit, they do not always agree on which is the lesser of two evils, or which is the best of two goods.

Indeed, even within each of the harms, there often is no agreed upon ranking, e.g., although both deafness and the loss of the use of one's legs are disabilities, it is rational for a person to rank either one of them as worse. Even if we are concerned with death and a specified duration and intensity of pain and suffering, it may be that there is no unique way to rank which is worse. Of course, sometimes it will be clear that one harm is worse, e.g., dying is worse than having the painful injection necessary to save one's life, which is why we sometimes talk of choosing the lesser evil. We all know, however, that rational people often disagree about which harm is worse, even though some philosophers, e.g., the Utilitarians, claim that all harms can be considered to be forms of pain, and talk as if there is always a correct way to rank all harms (6). This is one reason that people rightly pay so little attention to philosophers.

Medicine is concerned primarily with postponing, preventing, and relieving harms. Death, pain, and disability are the primary harms that accompany diseases and injuries. If a person's condition does not involve at least a significantly increased risk of death, pain, disability, or loss of freedom, e.g., allergies, or loss of pleasure, it should not even be counted as a disease. Abnormality, by itself, without any increased risk of suffering any of these harms should never be classified as a disease. Further, nothing should be classified as a disease if the increased risk of suffering any harm is solely because of the hostility of others. Homosexuality is no longer classified as a mental disorder, and hopefully transvestism also will soon lose that label (7). That these conditions make one more likely to be a victim of violence is an indication of problems in our society. It does not show that those who have these conditions are suffering from a disorder.

Now I shall go through each of the harms and briefly explain what I mean by them. The first harm is death and although many years ago I would have thought that it was quite clear what death meant, I am not so sure now. For the purpose of defining violence: causing death means causing a person to die in the biological sense, i.e., the organism permanently ceases to function, or causing permanent loss of consciousness. Causing pain includes not only causing physical pain, but also causing mental pain, e.g., anxiety, sadness, and anger. Causing disability involves causing both permanent disabilities, such as blindness, and temporary disabilities, such as broken arms or legs, etc. It also includes causing mental disabilities, such as phobias or addictions. Causing loss of freedom includes taking away a woman's freedom to do as she wants, e.g., by locking her up or physically restraining her. It also includes touching or administering medications without consent. Taking away a woman's pleasure involves interfering with her enjoyment, e.g., turning off her radio or television, or making such a commotion that she cannot hear it.

Violence involves intentionally or knowingly causing or trying to cause these harms. A man does not commit an act of violence if he is not paying any attention to where he is going and accidentally knocks you down. If he should have been paying more attention to where he was walking, he would be subject to moral censure, but he would not have committed an act of violence. Of course, you may be harmed as much as if he did knock you down intentionally, but the amount of harm suffered does not determine whether or not someone has acted violently. Violence as a way of causing harm is also different from causing harm by breaking a promise or deception. These ways of causing harm are also morally unacceptable, but they are not acts of violence. Causing harm is an act of violence only if one directly and knowingly causes the harm.

Violence never has the consent of the person who is harmed. One cannot consent to violence. This means that no matter what a doctor does to a patient, e.g., amputating his leg, it does not count as an act of violence if the patient has given his consent. Even if a doctor deceives a patient into consenting to an operation, which involves causing the patient to suffer some harm, this does not count as violence, whether or not the intent is to benefit the patient. Of course, such deception, except in very unusual circumstances, is still morally unacceptable, but it is not violence. Whether it counts as violence to operate on an unconscious patient from whom a doctor cannot obtain the consent, is a questionable matter, but if a doctor expects that the patient would have given consent, it seems not to count as violence. It is clearly an act of violence to operate on a patient contrary to his/her consent, even if he/she is not competent to give valid consent or a valid refusal, and the operation is necessary to save his/her life.

This example shows that violence is not always unjustified. As directly causing harm, however, violence always needs to be justified; but often, especially in emergency room settings, it can be justified. Somewhat surprisingly, the fact that violence needs to be justified limits who can act violently. Only moral agents, i.e., those whom we regard as at least partially responsible for their actions, can act violently (8). Animals, although they can certainly be violent in the secondary sense, cannot commit acts of violence. Animals do not have to justify anything they do, so if violence always needs to be justified, then animals cannot act violently. Whether one can commit an act of violence against animals is a more difficult question. Certainly one can be cruel to animals, but that is not quite the same. I think that the way that one answers this question depends on whether one regards animals as enjoying the protection of morality.

Putting all of these characteristics together I propose the following definition of violence. *Violence is an intentional or knowing attempt by a moral agent to directly cause harm, i.e., death, pain, disability, loss of freedom or pleasure, to someone who is protected by morality without the consent of that person.*

Given this analysis, it is clear that most interactions between health care workers and patients do not involve violence. Most patients do not intentionally or knowingly attempt to inflict harm on a physician, nurse, or other health care worker. This is not to deny that they may sometimes feel like committing an act of violence, but normally their fear or good sense prevents them from doing so. Although doctors, nurses, and technicians, often intentionally or knowingly inflict harm on patients, for the patient's own good, it is almost always with the consent of the patient, and so does not count as violence. Violence is not a normal feature of medical care; its occurrence is always a cause for concern and everyone in the health care field would prefer not to have to deal with it.

Patients do act violently toward health care workers and vice versa, however. Sometimes, of course, patients act violently in the secondary sense of the term; they do not intend or know that they are likely to harm someone, but their violent actions may significantly increase the chances of another patient or a health care worker being harmed. Sometimes this is a temporary condition, but it may also be permanent. Some patients do act violently in the primary sense, however; they intend to harm another patient or a health care worker, perhaps because they view that person as an obstacle to their getting out of the hospital, or simply to express hostility or rage. Such patients may be violent in both senses of the term.

Violence often begets violence. If a patient is violent in either sense of the term, responding violently toward that patient is often justified, i.e., either by temporarily disabling him with a drug, or by depriving him of his freedom with physical restraints. One can justify reacting to violence with violence if one intends to prevent harm to other people, either patients or health care workers. One cannot justify using more violence (i.e., the infliction of greater harm) than is necessary to prevent the harm that the patient's act of violence would cause. It is understandable that a health care worker confronted with a violent patient may be tempted to use more violence than is necessary, but that does not justify unnecessary violence. Although violence can be justified as a response to violence, whether a violent response is justified depends on the facts of the particular case. Among the most important of these facts is whether there is an alternative way of responding which would be equally effective by preventing that violence, but which involves less or no violence. Generally, it is morally unacceptable to use violence to control violence unless no alternative is available.

The paradigm of a violent patient is one who attempts to physically harm a health care worker. This can be done by biting, hitting, kicking pushing, shoving, etc. This can also be accomplished by squirting one's blood at a health care worker, if one is HIV positive, or even by breathing on a health care worker if one has tuberculosis. Any intentional or knowing attempt to cause harm to someone without their consent counts as an act of violence. An abusive patient who screams and threatens is also violent, although it is not always clear whether this is violence in the primary or the secondary sense, i.e., whether patients know that they are causing harm to someone, either a health care worker or other patients, or are simply out of control and do not realize what they are doing to others.

The paradigm of a violent health care worker is one who reacts to the violence of a patient inappropriately, i.e., responding to a patient's violence with more violence than is necessary to prevent harm to others. But not all violence by health care workers is inappropriate. Sometimes it is both necessary and justified to respond to the violence of a patient with violence, e.g., to sedate or physically restrain a violent patient. Violence is appropriate only to prevent violence, not to punish it, in a hospital setting.

Hospitals are not equipped to deal with violence by patients. It is difficult enough to practice good medicine with patients who merely do not understand or appreciate what they must do to participate in their own care. Patients violently resisting appropriate treatment present a new level of difficulty. Can hospitals refuse to treat violent patients, that is, those who are violent in the primary sense? It is plausible that hospitals can have a public policy which states that they will not provide non-emergency treatment to patients who are violent. I am not talking about patients who do not know what they are doing, I am talking about those who intentionally or knowingly attempt to harm other patients or health care workers.

There are two other problems concerning violence that deserve some attention. The first concerns the duty of health care workers when they discover that the injuries suffered by a patient are the result of violence by someone else. If the victim is a child, the law requires that the matter be reported to the appropriate authorities. The advantages of having this law seem to outweigh its disadvantages, for we know that if people were given discretion in reporting, there would be significant under-reporting and more children would suffer than they would when reporting is required. If the victim were an adult, usually a woman who is being physically abused by the man she lives with, the question would be more difficult. A woman can, if she is provided with adequate protection, report the violence herself. But adequate protection often cannot be provided, so not only will a woman risk additional harm if she reports the violence herself, she will also risk additional harm if someone else reports the violence.

Given these facts, it seems that there should not be a requirement to report violence against adults. However, that does not mean that the problem should be ignored, as now seems to be the normal procedure. There should be a requirement that victims of violence be informed of their legal and other options. Counseling should be made available, as should as much protection as possible. Everything should be done to enable the victim herself to report the violence and to escape the violent situation. Sometimes, even completely adequate protection, a rarity, will not be sufficient to persuade the victim to report the violence. There may be children involved, and emotional or financial dependency. Until these problems are solved in the society at large, it is pointless to expect hospitals and health care workers to solve them.

The second problem concerns the duty of health care workers when they discover that a patient is planning to commit an act of violence. If the patient is suffering from a mental disorder, the problem is already well known to psychiatrists: when is it appropriate to commit a patient who is a danger to self or others? There has been considerable discussion of this topic, and there seems to be no best solution. Statistics show that the reliability of predictions of violence is poor and, even if it were better, one still would have to choose between committing many patients who would not be violent even if not committed and not committing some who will be violent if not committed. Thus states go back and forth with legislation which makes requirements for commitment stronger, e.g., an overt violent act must have been attempted, and making them weaker, e.g., verbal threats are sufficient for commitment. This problem is one for which there is not only no best practical solution, but there is also no best philosophical solution.

When dealing with a violent patient who is not suffering from a mental disorder, matters seem to be simpler. Preventive detection is subject to so many abuses that, without adequate safeguards and predictions of violence that are close to 100% reliable, it does not seem a plausible option. Hospitals are not prisons, and health care workers are not police, not even paid informers for the police. Of course, if there is a specific threat against a particular person, it is permissible to report this to the police, and even to warn the person threatened, but it is not clear that this will do much good. The debate about the Tarasoff decision shows that there are strong arguments both for and against requiring warning specific victims about threats from patients. The statistics are not good and it is not clear which policy will result in a lessening of the harm suffered. Rational people can advocate different policies, but there does not seem to be a best policy, either practically or philosophically.

Of course, these observations do not solve the problem of violence in medicine; however, they do confirm that more needs to be learned about the consequences of different policies. It is also clear that many of the problems of violence must be solved by those outside the health care field; that one cannot expect hospitals and the health care field to solve problems that are part of the larger society. As a philosopher I am not competent to provide ways of dealing with violence, I can simply clarify what counts as violence and identify some of the ethical issues that violent patients and victims of violence present to hospitals and health care workers.

References

1. See Gert B. Justifying violence. J Philos 1969; 66(19): 616–628, for a more extended discussion of violence. I shall, in what follows, simply talk of "causing harm" not of "trying to cause harm," but what I say applies equally to both.
2. See Gert B. Morality—A new justification of the moral rules. New York: Oxford University Press, 1988, especially Chapter 3.
3. By this I mean "avoidable death" which is generally premature death. I do not deny that many people, at some point in their lives, come to regard death as the lesser evil, or as Hobbes says, the pains of life can be so great that death is numbered among the goods.
4. See Gert B. Rationality human nature, and lists. Ethics 1990; 100(2):279–300; and Defending irrationality and lists. Ethics 1993; 103(2):329–336.
5. See Gert B. Morality—A new justification of the moral rules. New York: Oxford University Press, 1988, Chapter 2.
6. This becomes even clearer when one realizes that most often the degree of risk of harm is involved. Although everyone would rank death as worse than being deprived of some pleasure, it is clear that this same ranking would not hold universally if it were a continuing pleasure, e.g., drinking, and it was only a slight increase in the risk of death, e.g., less than 1%.
7. See Gert B. A sex caused inconsistency in DSM-III-R: the definition of mental disorder and the definition of paraphilias. J Med Philos 1992; 17(2):155–171.
8. See Gert HJ. Rights and rights violators: a new approach to the nature of rights. J Philos 1990; 87(12):688–694.

The Legal Obligation to Prevent Harm:

The Tarasoff Rule and Its Limits

Daniel Givelber, J.D.

Discussing legal arrangements that may influence a physician's response to violence, either by a patient or against a patient, ought to start with an understanding that the existence of any legal background at all is itself worthy of note. Thirty years ago my guess is that this symposium would not have been held—but, if it had been, the legal background would have been disposed of quickly. Other than an obligation to report gunshot wounds, the "law" on violence probably would have been reduced to matters of evidence in court, such as whether a physician could be compelled to testify in court about a conversation with a patient. The notion that health care providers were legally obligated to "do something" about violence simply did not exist. Thirty years ago, whatever the moral case for intervention, there was no legal obligation to notify the child welfare authorities if a child was brought to the emergency room with suspicious injuries. Or a physician might have suspected incest, but what he did with those suspicions was a matter of institutional culture and personal and professional ethics, as opposed to formal, affirmative legal compulsion. It was neither a tort nor a crime to fail to report suspicions of abuse. And the idea that even if a physician decided to intervene he would ever have to do more than simply report a problem to public authorities was preposterous.

Things have changed during the past 30 years. Now, mandated reporting of suspected child abuse—physical, sexual, and possibly even psychological—is the law virtually everywhere (1). There is law, particularly legislation and administrative decisions, relating to the health provider's obligation to intervene with respect to the sexual partners of an HIV-positive patient. Many states have such laws, although they focus on protecting the physician from liability regardless of what he or she does rather than on requiring a particular approach (2). There is also law, judicial decisions, rather than legislation, that requires at least one subset of health care providers, those dealing with mental illness, to intervene to protect third parties from the potential violence of the patient.

Despite these legal developments, if you are confronting a difficult ethical dilemma triggered by concern for the safety of your patient or others, the chances are that you will still be facing a difficult ethical dilemma once you have learned what the law has to say. One reason for this is the legal ambivalence toward our obligation to help strangers.

A classic law school hypothesis dealing with the duty to rescue is the child in the baby carriage rolling ever so gently into oncoming traffic. If there is a group of people at the intersection, any one of whom could easily reach out and stop the carriage before it rolls in front of oncoming taxicabs, but no one bothers to stop it, traditional doctrine is that none of these people has committed either a crime or a tort. Only someone who has had a preexisting relationship with the child (for example, a babysitter or a parent) or who undertook to help the child or had a role in setting the carriage rolling could be held to be tortious or criminal for the failure to rescue. There is no obligation, either criminal or civil, to rescue strangers.

The situation might change if one of those who had no preexisting relationship reached out for the carriage, saying, "Don't worry—I'll stop it from going into the street." Particularly if the re-

---

Professor of Law, Northeastern University Law School.

sult of the volunteer's statement was to satisfy other potential rescuers that their efforts were not needed, this behavior would be viewed as a voluntary undertaking generating a legal obligation to exercise reasonable care to protect the child. If the volunteer failed to take such care, the volunteer would be liable.

If, instead of observing the carriage rolling toward Fifth Avenue, a passerby was simply told that the carriage was in danger and the passerby did nothing, there would be no liability. If strangers have no obligation to act when they see disasters unfolding, they have no obligation when they simply hear about it. Nor does the situation change if the person who tells them about the rolling carriage also says, "And I want you to know that I am the very person who set that carriage in motion!"

If I do not need to reach out my hand to stop a baby carriage about to roll into oncoming traffic, I do not need to reach out to stop the carriage simply because I see or have been told that someone deliberately pushed the carriage toward the oncoming traffic. A legal system that does not insist on the duty of easy rescue is also a legal system that does not require individual citizens either to take steps to prevent crime or to attempt ameliorating its effects.

Suppose that the passerby looks up and sees that it is someone he or she knows who has pushed the baby carriage toward the street. Suppose it is a neighbor? No liability. A spouse? Again, no liability. A client (assuming the passerby is a lawyer or an architect or a CPA)? No liability. A patient? Well, if an individual is not responsible for the baby in the careening carriage if the carriage has been set rolling by a neighbor or a loved one, why should that individual be responsible simply because she has a professional medical relationship with the person who sets the carriage in motion? Although the precise case has not arisen, the answer is pretty clearly that a physician would have no legal responsibility.

Now suppose that instead of seeing the baby carriage rolling toward the street, our passerby meets someone who says, "I just want you to know that I intend to push that baby carriage into the street in front of oncoming traffic." Does this change anything? By extension, it seems that if one has no obligation to a stranger who is in actual peril, one should not have an obligation to that same stranger who is only in potential peril. This holds true for lawyers and philosophers and for most physicians, but it no longer holds true for those providing mental health services. If the person who threatens to harm a stranger is a patient, and the psychiatrist believes or reasonably should believe that the patient is likely to carry out the threat, then the psychiatrist has a legal obligation to take reasonable steps to protect the potential victim.

If I see a child who is bruised as he or she leaves her house, I am not required, either as lawyer or citizen, to call the child welfare authorities. If I am told by a friend that he is HIV positive and intends to continue practicing unprotected sex with partners who do not know his status, I have no obligation to do anything about that. This is true even if I know those sexual partners and could easily advise them of their risk. And I need not do anything if I am sitting at a bar and the person on the next stool convinces me that he is about to go home and shoot his spouse. So why should mental health providers have this obligation?

In this society, we do not have a general affirmative obligation to protect strangers. As long as we are not engaged in conduct that can be viewed as causing harm to another, the fact that we withhold our aid is a matter for our conscience. Legally, we are not held accountable. A major justification for this rule has been that any other rule poses definitional problems that threaten individual autonomy and liberty. If I have a legal obligation to stop the carriage from rolling into the street, do I also have a legal obligation to give money to the homeless person? To use Macauley's famous example, does a surgeon have to travel from Calcutta to Meerut if she is the only person who can successfully perform an operation? The difficulty of articulating meaningful criteria that would permit each of us to know when we must and when we need not try to help a stranger, coupled with a profound commitment to the value of freedom of action, as well as a reluctance of judges to make major changes in basic social arrangements, gives us tort and criminal law that requires us to behave reasonably when we act but does not require us to be good Samaritans. Thus, perhaps for no greater social reason than that it interferes with my liberty, I cannot be compelled to aid others.

If that is the general rule of tort law—and it remains the rule—then one would think that there would be an even more compelling argument against an obligation to aid a stranger when the good Samaritan action would be inconsistent with a preexisting obligation to someone else. Thus, if you, as a physician, are treating a patient or I, as a lawyer, am representing a client, the assumption is that we need to assure our client confidentiality. It is essential to create the atmosphere of confidentiality and to generate the flow of information necessary to maximize the effectiveness of what we can do to help the patient or client. Because I already have a relationship with my client, and there is a positive value in maintaining that relationship so that I can provide my socially useful service, and confidentiality helps to maintain that relationship, then it is clear that I would have no obligation to aid a stranger whose interests my client is threatening. As a lawyer, I have no such obligation. But when we change professions and start talking about people providing health care, particularly mental health care, the logic no longer holds. Rather than being an additional reason for there being no duty to intervene, the doctor–patient relationship becomes the very justification for imposing a duty to care for a stranger.

The Tarasoff Case

The case which produced this remarkable development is *Tarasoff v Board of Regents of the University of California* (3). According to the allegations made by the plaintiff (and treated by the court as true for purposes of the decision), a graduate student named Prosenjit Poddar sought the services of the psychiatric clinic at the University of California, Berkeley, where he was seen by a psychologist. The therapist concluded that Poddar represented a threat to a young woman, Tatiana Tarasoff, and sought to have Poddar committed involuntarily. Because the psychologist lacked the legal authority under existing California law to order the involuntary commitment, he requested the campus police (who had such authority) to make the commitment. The police briefly detained Poddar before concluding that he was not a danger and releasing him. Tarasoff was not warned of what had transpired; 2 months later, Poddar killed her.

The California Supreme Court ruled that when a therapist knows, or should reasonably have known, that a patient poses a threat of physical violence to an identifiable third party, the therapist owes an obligation to exercise reasonable care to protect that third party. This exercise of reasonable care may, but need not, take the form of warning the victim.

Given this legal rule, if the plaintiff in the *Tarasoff* case could prove the facts alleged, and the jury agreed that simply notifying the campus police did not represent a reasonable response to the threat, the plaintiff would win and the psychologist and the clinic would be liable. Had the rule stopped here, we probably would not be discussing it. The rule goes further, requiring action even in those situations in which the therapist, in good faith, does not believe that the patient poses a threat. If the plaintiff establishes that, under existing professional standards, the therapist *should* have determined that the patient was a threat to an identifiable victim, the therapist can be liable for failing to take reasonable steps to protect the victim. This rule—requiring action whenever the therapist should predict the patient will be violent—has been endorsed by most courts confronted with the question.

Implications of *Tarasoff*

From a legal point of view, the *Tarasoff* decision was and is an "odd duck," full of inconsistencies. It is unprincipled in the sense that neither the California courts nor those of any other state have viewed it as embodying a principle of universal application. What is it about the therapeutic relationship that gives a psychiatrist an obligation that no one else has? This rule has not been applied to health care providers other than therapists, at least with respect to predicting dangerous behavior. It has not been applied to the police or lawyers or any other professionals likely to come in contact with potentially dangerous individuals. Moreover, and even more curiously, California, like most other states, has refused to apply the Tarasoff rule to a therapist's failure to take steps to warn others that a loved one may be suicidal, or to situations where it is predictable that the patient will injure some unidentifiable person. And it is still the law everywhere that none of us need reach out and stop the baby carriage as it rolls toward traffic.

The reasoning behind this distinction is a bit opaque. Several possibilities for treating psychiatry as a special case suggest themselves:

- The psychiatrist can predict who is going to be dangerous and it is the power to predict which differentiates the psychiatrist from everyone else.
- The psychiatrist learns the patient's innermost thoughts, and it is this access to what the patient is "really thinking" which gives the psychiatrist the unique obligation.
- The psychiatrist is empowered by the state to deprive a person of freedom for a brief period of time on the grounds of future dangerousness. The holder of this sort of power ought to use it to protect innocents.
- Professional ethics already require action in such situations. We are now putting the power of the law behind the existing ethical mandate.

Each of these is problematic. (a) There is no persuasive evidence that mental health practitioners are good at predicting future violence. Moreover, if making professional judgments about whether someone is likely to be dangerous is the crux of the rule, then law enforcement personnel and judges should have the same obligation. Federal law permits preventive detention based on judicial determinations of future dangerousness. (b) The rule ought to apply to anyone with access to private thoughts and fantasies. Yet it does not apply to a lawyer, a lover, a bartender. (c) The police, who have the specific authority to commit individuals involuntarily, are specifically exempted from responsibility under the *Tarasoff* ruling. (d) We do not convert all moral obligations into legal ones. The professional norm has been one that permitted intervention without requiring it.

Whether *Tarasoff* will apply more broadly, of course, depends on which of these rationales is the one a court chooses to explain the decision.

Invoking the *Tarasoff* Decision: Disease Transmission

When the threat is not physical violence but the transmission of a disease, a duty to strangers may arise. Physicians who are not responsible for protecting third parties against the possibility of violence from a patient may have an obligation either to protect third parties from their patient's disease or, at least, to refrain from anything that might enhance the risk to third parties. *Tarasoff* has been invoked in these cases.

In *Reisner v Regents of the University of California* (4), the defendant physician allegedly unwittingly transfused a 12-year-old girl during surgery with blood contaminated with HIV antibodies. Although the physicians learned of the contamination on the following day, they never informed the patient or her parents. Three years later, the patient began dating Reisner, the plaintiff, and the two became intimate, which resulted in Reisner contracting HIV. The court upheld Reisner's suit against the physicians based on the theory that the physicians had a duty to their patient to notify her or her parents that she had received contaminated blood and that, had they done so, she or they could have informed Reisner before he became exposed to the virus. As the court notes, the case involves enormous questions of causation. Would the patient have notified the boyfriend had she learned directly that she was infected with the virus?

What about the situation in which a disease is not communicable through human interaction but its existence in one member of a family creates a concern that others might be at risk? *Bradshaw v Daniel* (5), a Tennessee case, presented the question of whether a physician whose patient suffers from Rocky Mountain spotted fever has a legal obligation to advise the patient's wife about the diagnosis and to alert her to the risk of contracting the disease. As the court understood the transmission pattern, the wife's risk was not that her husband could communicate the disease to her but that she had also been exposed to the infected ticks. The physician argued that the lack of a patient-physician relationship with the wife, coupled with the fact that Rocky Mountain spotted fever is not a contagious disease, means that even if he had a medical obligation to inform the family of the diagnosis, he did not have a legal one. Relying on *Tarasoff,* the court held that the physicians had a legal, as well as a medical obligation to warn identifiable persons in the patient's immediate family of risks to them "emanating" from the patient's illness.

In *Reisner,* the physician (by definition) did something unreasonable, vis-à-vis the patient, by failing to communicate to her and her family that she had been transfused with HIV-infected blood. It was the consequence of this malpractice against the patient that caused the forseeable injury to a third party. In *Bradshaw,* the plaintiff asserted that the physician negligently failed to diagnose Rocky Mounted spotted fever in her husband. Although as a technical matter the court's decision did not depend on the physician's mistreatment of his patient, the physician's failure to treat the husband adequately facilitated the ruling that the physician also failed in his obligation to the wife. In *Tarasoff,* it is more than arguable that the treating psychologist did absolutely everything he could have done for the patient. Yet he is still held liable.

Conclusion

There is no ready answer to the question of why psychiatrists and mental health workers have a responsibility that other health professionals may not have. The distinction probably rests on a combination of factors—the difficulty of predicting violent behavior, the desperate wish that we could successfully predict it, the legal and professional authority to make such predictions enjoyed by the mental health profession, the desire to attribute powers to mental health professionals that they do not have. Because the exact reason for the *Tarasoff* rule is so difficult to identify and because the courts seem happy with the literal rule and are not eager to expand it, courts are not likely to ask other health care professionals to make the kinds of judgments about future human behavior which they demand of mental health professionals.

Although it is possible for a court at any particular point to determine that protecting third parties trumps the values inherent in maintaining patient privacy and respecting the confidentiality of patient communication, courts to date have resisted the open invitation provided by the *Tarasoff* doctrine. It is out there, rich with potential, but not yet consistently carried to where logic suggests that it might go.

References

1. Kalichman S. Mandated reporting of suspected child abuse: ethics, law and policy. Washington, D.C.: American Psychological Association, 1993.
2. Bayer R, Toomey K. HIV prevention and the two faces of partner notification. Am J Public Health 1992; 82: 1158–1164.
3. Tarasoff v Board of Regents of the University of California, 551 P2d 334 (1976).
4. Reisner v Board of Regents of the University of California, 31 Cal App 4th 1195 (1995).
5. Bradshaw v Daniel, 854 SW2d 865 (Tenn, 1993).

Autonomy and Violence

Bat-Ami Bar On

The ideas of violence and victimization are bound together so much that at first it seems hard to imagine victimless violence. Using war as my example, I asked myself: "How could violence have no victims and not have them as more than just a matter of accident?" Yet, once I began thinking about the possibility of victimless violence, I realized that the violence of war is complicated. Not only are people who accidentally escape bullets or bombs not physically victimized by a war's violence but also not everyone physically or psychologically hurt by a war's violence is considered one of its victims. Even soldiers who are killed, maimed, or wounded during a war are not considered to be its victims as long as the war is conducted with enough attention to the rules that are expected to limit what warriors do to each other, rules like those that the Church devised for medieval knights or those specified in the Geneva Convention.

A fairly standard way to explain why a soldier is not a victim of a war's violence, and one built into various war conventions and treaties, is an appeal to the equality of soldiers, an equality they share exactly because they are soldiers. According to Michael Walzer (1), soldiers not only are equal but perceive themselves as such, especially when in one way or another they are compelled into servitude as soldiers. Abstracting from soldiers' memoirs, he describes a soldier's growing comprehension of the equality of soldiers as follows:

> [W]e know that [the soldiers on the other side] are not responsible [for the war]. . . . [T]he enemy soldier, though his war may be criminal, is nevertheless as blameless as oneself. Armed he is the enemy; but he isn't my enemy in any specific sense; the war itself isn't a relation between persons but between political entities and their human instruments. These human instruments are not comrades-in-arms in the old style . . . they are "poor sods, just like me," trapped in a war they did not make. I find in them my moral equals. (p 36)

One could interpret Walzer as describing an equality of soldiers that is based on their similar victimization, and he himself refers to his soldiers as victims. He does so in a way that suggests that the victimization, although immediately caused by the war, is a function of what compels people into the soldier's servitude to the state, whether it is the state's ability to coersively demand such servitude or economic conditions. For him, soldiers recognize this when they come to see themselves as equal. It follows for them to have their own sense of what counts as victimization by the violence of war. Walzer tries to capture this soldier's sense of what counts as victimization with five examples of soldiers who refrain from attacking because what they saw was not an enemy but someone who looked funny, taking a bath, holding up his pants, reveling in the sun, smoking a cigarette. Walzer writes,

> War is not in fact [a soldier's] enterprise, but rather surviving this battle, avoiding the next. Mostly, he hides, is frightened, doesn't fire, prays for a minor wound, a voyage home, a long rest. And when we see him at rest, we assume that he is thinking of home and peace, as we would be. If that is so, how can it be justified to kill him? Yet it is justified, as most of the soldiers in the five stories understand. Their refusals seem, even to them, to fly in the face of military duty. Rooted in a moral recognition, they are nevertheless more passionate than principled decisions. They are acts of kindness, and insofar as they entail any danger at all or lower minutely the odds of victory later, they may be likened to supererogatory acts. Not that they involve doing more than is morally required; they involve doing less than is permitted. (pp 142–143)

Walzer can liken the five soldiers' choice not to attack to supererogatory acts because he interprets their actions as residing in a reestablishment of a sense of common humanity and, conse-

---

From the Department of Philosophy, State University of New York at Binghamton, Binghamton, NY 13902-6000. Dr. Bar On is Associate Professor of Philosophy.

quently, a reassertion of equal rights, specifically equality of the right to life. But an equal right to life can be asserted in ways that lead back to a victimless violence, not away from it.

The Right To Be Violent

For Thomas Hobbes (2), an equal right to life is the basis for a right to violent self-defense, a right that cannot be transferred as part of the mutual transfer of rights that constitutes the social contract. Hobbes says,

> [T]here be some rights that no man can be understood by any words, or other signs, to have abandoned, or transferred. As first a man cannot lay down to resisting them, that assault him by force, to take away his life; because he cannot be understood to ayme thereby, at any Good to himself. (p 192)

In the picture that Hobbes paints, the would-be resistor has an absolute right to resist and resist violently anything that could count as a possible attack. But, when the resistor acts on this right, the person suspected of the possible attack becomes a resistor too because the right to life is threatened and the right to resist is activated. The result is the "state of war" that Hobbes' social contract is supposed to transform. Although Hobbes' "state of war" may seem to be a state in which nothing but victimization is possible, in a deep sense, it is a victimless state because it is a state without justice.

Hobbes, more clearly than Walzer, shows that the connection between violence and victimization is mediated by ethicopolitical rules, guidelines, or principles that result from the kind of social transformation that is put into place by the social contract. Where socially located ethicopolitical rules, guidelines, or principles are enforced, one can distinguish between victimizing and victimless violence. This is a point that John Locke makes even clearer when he discusses punishment. For him, punishment is not victimizing as long as it is proportionate to the transgression.

The Pleasure of Violence

Nietzsche (3) provides a genealogy of the ethicopolitical mediation of the relationship between violence and victimization. According to him, a certain kind of noble ethicopolitics permits a victimless violence quite different from the kind imagined by Walzer, Hobbes, and Locke, at least toward outsiders, people with whom one has no ethicopolitical understandings of any sort. As Nietzsche describes it, violence against outsiders is not only victimless but also pleasurable. The pleasure derives from the uninhibited articulation of strength. He explains that "pleasure is induced by his being able to exercise his power freely upon one who is powerless, by the pleasure of . . . rape." (p 196) He adds that:

> To behold suffering gives pleasure, but to cause another to suffer affords an even greater pleasure. This severe statement expresses an old, powerful, human, all to human sentiment. . . . There is no feast without cruelty, as men's entire history attests. Punishment, too, has its festive features. (p 198)

According to Nietzsche, the pleasure that is derived from suffering suggests that the distinction between victimizing and victimless violence is a rational cover for something that is irrational. This disruption neutralizes the normative force that suffering is otherwise expected to exert and reverses the ordinary condemnation of pleasure derived from suffering.

Nietszche's disruption of the ethicopolitical mediation between violence and victimization and his neutralization of the normative force of suffering are quite disturbing, especially because they confront one with the pleasures of cruelty. Still, Nietzsche's ideas are useful, enabling one to make better sense of phenomena like torture, for example. I find them even more useful, however, exactly because they disrupt and disturb. They cast suspicion even on one's motivation in responding to suffering, on the emotions that one may have as one reads about torture in the Argentine junta's detention centers, or watches documentaries from the Holocaust, or witnesses the aftermath of a bombing in Bosnia or Iraq or the aftermath of smaller events, such as a shooting, a rape, or a battering that might have taken place in one's neighborhood. Straightforwardly, Nietzsche claims:

> Perhaps it is even legitimate to allow the possibility that pleasure in cruelty is not really extinct today; only, given our greater delicacy, that pleasure has had to undergo a certain sublimation and subtilization to be translated into imaginative and psychological terms in order to pass muster before even the tenderest hypocritical conscience. ("Tragic empathy" is one such term; . . . .) (p 200).

Listening to Accounts of Suffering

I would like to expand on the issue of listening to suffering, because that is what one does when listening to someone who has been victimized. Those who have done this kind of listening are aware of how hard it is to do. It is a hardship that the literature on trauma describes as a function of the destruction of language. As Judith Herman (4), a therapist specializing in incest, notes:

> The ordinary response to atrocities is to banish them from consciousness. Certain violations of the social compact are too terrible to utter aloud; this is the meaning of the word *unspeakable*. (p 1)

According to Lawrence Langer (5), it is not the terror of an atrocity alone that produces unspeakability. Rather, unspeakability is produced by an estrangement between the victim/survivor who has or continues to suffer and the person who listens to her or him. Because of their different experiences, the two do not share the same normative world. The victim/survivor's world is qualitatively different from the ordinary normative world of the listener. Not sharing a normative world, their communication is problematic. To the victim/survivor this is obvious and Langer, who has listened to victims/survivors of the Holocaust, says that when a victim/survivor asks:

> "Do you understand what I am trying to tell you?"–that witness confirms the vast imaginative space separating what he or she has endured from our capacity to absorb it. (p 19)

One may refuse to enter the vast imaginative space that separates the victim/survivor from the listener. Herman believes that one reason the entrance is refused is because the victim/survivor's ethicopolitical demands are weighty, whereas those of the perpetrator are easier to live with. After suggesting that a witness is forced to take sides and choose between that of the victim/survivor and that of the perpetrator she writes:

> It is very tempting to take the side of the perpetrator. All the perpetrator asks is that the bystander do nothing. He appeals to the universal desire to see, hear, and speak no evil. The victim, on the contrary, asks the bystander to share the burden of pain. The victim demands action, engagement, and remembering. (pp 7–8)

Siding with the perpetrator is, for Herman (4) made easier by social inequities and isolation. She claims:

> After every atrocity one can expect to hear the same predictable apologies: it never happened; the victim lies; the victim exaggerates; the victim brought it upon herself; and in any case it is time to forget the past and move on. The more powerful the perpetrator, the greater is his prerogative to name and define reality, and the more completely his arguments prevail. The perpetrator's arguments prove irresistible when the bystander faces them in isolation. Without a supportive social environment, the bystander usually succumbs to the temptation to look the other way. This is true even when the victim is an idealized and valued member of society [like the heroic soldier]. . . . When the victim is already devalued (a woman, a child), she may find that the most traumatic events of her life take place outside the realm of socially validated reality. (p 8)

Some may be willing to take on the burden of listening. According to Langer (5), this requires repetitive listening that at one and the same time particularizes and familiarizes because, through the repetitions, the victims/survivors become specific people from specific places and times, all named, people with a past full of family and friends, with habits, wants, fancies, and daily sorrows who, nonetheless, are the victim/survivors of traumatizing events. Langer cites Magda F., victim/survivor of the Holocaust saying:

> [T]o understand us, somebody has to go through with it. Because nobody, but nobody fully understands us. You can't. No [matter] how much sympathy you give me when I'm talking here, or you understand . . . you're trying to understand me, I know, but I don't think you could. I don't think so. (p xiv)

Langer responds by saying,

> A statement like "to understand, you have to go through with it," however authentic its inspiration, underestimates the sympathetic power of the imagination. Perhaps it is time to grant that power the role it deserves. (p xv)

I do not think that one could grant "the sympathetic power of the imagination" the role that Langer believes that it deserves and not take to heart Nietzsche's arguments about the pleasure of violence. One should concern oneself with the possibility that one's tender conscience is hypocritical. This is a possibility that Herman does not pursue. And Langer, who at various places turns to Nietzsche for analytic tools, only hints at it as he discusses the formation of a deep memory of the traumatizing events and its rupturing of the memory that is formed after those events, the anguish of remembering that even in Auschwitz one yearned for life, the ways one is humiliated and tainted by one's memories because the memories crush while there is no consolation and there can be none because there is no heroism in surviving.

Langer demands that his readers attend self-reflexively as listeners to the victim/survivor. The demand is incorporated into Langer's writing because his readers are also readers of victims/survivors' testimonies. An excerpt of an interview with Moses S. which exemplifies how even momentary escape of acting unreflexively out of a tender consciousness can discount the victims/survivors' experiences.

> *Moses S.:* All right. A few weeks later, the English people came in and bombed the concentration camp [Mauthausen]. And, I said, "Yankel, get up, get up, it's no good lying here, you'll be a piece of *gornish* [nothing at all]." So, we got up, and we found a hand from the bombing.
> *Interviewer:* A hand grenade?
> *Moses S.:* No, a hand [At this point, another voice, presumably a family member, interjects: "A human hand"].
> *Moses S.:* A human hand.
> *Interviewer:* Oh, a human hand.

> *Moses S.:* Five of us. Divided. And we were eating it. And somebody died, we cut out a piece—we were eating . . .
> *Moses S.'s wife:* Excuse me, I think we have to finish. Too much already.
> *Moses S.:* Human flesh. (p 117)

Moses S.'s talk of cannibalism is talk of a violation of a taboo that is in the West an important marking of the civilized. When his interviewer and wife misunderstand him or try to stop him from speaking, they act as the guardians of the culture. They also guard themselves against a confrontation with their own hypocrisy when pushed beyond the initial horror at the cannibalism itself. One can hardly avoid a confrontation of this kind when reading Tadeusz Borowski (6), who like Moses S., is also a victim/survivor and describes the sensual pleasures that Auschwitz inmates derived from a kind of metaphoric cannibalism. At the very beginning of his story Borowski makes it clear that sensual pleasure was possible for an Auschwitz inmate. The pleasure was derived from items that were taken from transportees who were gassed. He writes,

> Several of us sit on the top bunk, our legs dangling over the edge. We slice the neat loaves of crisp, crunchy bread. It is a bit coarse to the taste, the kind that stays fresh for days. Sent all the way from Warsaw—only a week ago my mother held this white loaf in her hands . . . dear Lord, dear Lord. . . . We unwrap the bacon, the onion, we open a can of evaporated milk. Henri, the fat Frenchman, dreams aloud of the French wine brought by the transports from Strasbourg, Paris, Marseille. . . . Sweat streams down his body, "Listen, *mon ami,* next time we go up on the loading ramp, I'll bring you real champagne. You haven't tried it before, ah?" "No. But, you'll never be able to smuggle it through the gate, so stop teasing. Why not try and "organize" some shoes for me instead . . .?" "*Patience, patience*. When the new transports come, I'll bring all you want. . . ." "And what if there are no more "cremo" transports? . . . [D]ammit, they'll run out of people." . . . "Stop talking nonsense." Henri's serious fat face moves rhythmically, his mouth is full of sardines. . . ." "Stop talking nonsense," he repeats, swallowing with effort. "They can't run out of people, or we'll starve to death in this blasted camp. All of us live on what they bring." (pp 30–31)

After this beginning, the story describes the unloading of transports at Auschwitz. They are shown to be terribly horrifying because they are routine, hence, seemingly ordinary. So, for example, Borowski's says of the loading station for Birkenau, "This is where they load freight for Birkenau: supplies for the construction of the camp, and people for the gas chambers. Trucks drive around, load up lumber, cement, people—a regular daily routine." (pp 33–34) At the end, the story returns to the point it makes at the beginning about the sensual pleasure of inmates and the metaphoric cannibalism typical of Auschwitz. Borowski relates:

> It is almost over. The dead are being cleared off the ramp and piled into the last truck. The Canada men, weighed down under a load of bread, marmalade and sugar and smelling of perfume and fresh linen, line up to go. For several days the entire camp will live off this transport. For several days the entire camp will talk about "Sosnowiec-Bedzin." "Sosnowiec-Bedzin" was a good, rich transport. (p 49)

Ending the story as he began it, Borowski does not allow his readers or himself to escape by dwelling only on the atrocities of the Nazis and their collaborators. He seems to want his audience to dwell on a bizarre innocence reminiscent of the innocence of Nietzsche's nobles. One cannot enjoy it with a clear conscience. For Borowski, it is, nonetheless, an important kind of innocence because he finds his work with the transports nauseating, yet is alive and is among others who are still alive. Thus he says:

> I lie against the cool, kind metal [of the rails] and dream about returning to the camp, about my bunk, on which there is no mattress, about sleep among comrades who are not going to the gas tonight. Suddenly I see the camp as a haven of peace. It is true, others may be dying, but one is somehow still alive, one has enough food, enough strength to work . . . (p 48)

The structure of Borowski's story insists on its own terms. Its readers must be profoundly respectful of the victim/survivor as the narrator of his own story. According to Dori Laub (7), a psychoanalyst who specializes in trauma, this kind of profound respect is essential to the therapeutic process that makes the traumatizing event known. For Laub, this has radical implications. Telling the story of interviewing a woman victim/survivor of Auschwitz, he points out that what she said was disputed by historians as lacking in accuracy. He distinguishes between himself and the historians by saying of himself,

> My attempt as interviewer and as listener was precisely to respect—not to upset, not to trespass—the subtle difference between what the woman *knew* and what she *did not, or could not know*. It was only at the price of this respect, I felt, this respect of the constraints and of the boundaries of silence, that what the woman *did know* in a way that none of us did—what she came to testify about—could come forth and could receive a hearing. (p 61)

Laub is not advocating for listeners to be ignorant. What he asks the listener to do is to take the testimony not as an experiment that can be expected to replicate and be replicated, but as a different kind of knowledge. This is a knowledge of a traumatizing event that is unique to the victim/survivor and conveys much more than an unreliable kind of subjective truth. It is a knowledge that illuminates the traumatizing event in a particular way no matter what the facts are.

Laub grants the victim/survivor a strong kind of autonomy. To follow him one would have to resist one's inclination to separate oneself from others by thinking of them as victims/survivors in light of reasonable evidence one has that is independent of their own testimony. For Laub, especially when the person one encounters seems to be victimized and may be at risk of further and perhaps worse victimization, one needs to enter their vast imaginative space and try to comprehend their experience.

But should one follows Laub's advice? There is something maddening about this dilemma. If one goes along with Laub and grants others a strong autonomy, one may escape the charge of claiming to know better than someone else what their experience is like for them. At the same time, one may, thereby, deny the need to assume a different set of ethicopolitical responsibilities toward others, such as the responsibility to aid someone in need. Maddening as the dilemma is, however, I do not believe that one should rush to solve it abstractly.

Abstract or Particular Solutions

Nietzsche (3) says that:

> What makes people rebel against suffering is not really suffering but the senselessness of suffering; . . . . In order to negate and dispose of the possibility of any secret unwitnessed suffering, early man had to invent gods and a whole apparatus of intermediate spirits . . . Such were the inventions with which life, in those days, performed its perennial trick of justifying its "evil"; nowadays a different set of inventions would be needed, e.g. life as a riddle or an epistemological problem. (pp 200–201)

Like Nietzsche, I do not think that suffering can be eliminated. I also do not think that it should be naturalized or ethicopolitically neutralized by justifying theories like Nietzsche's. Nonetheless, I think that Nietzsche is right in suggesting that it is not suffering, as such, but meaningless suffering that moves one to try to make sense of it. To attempt to solve the abstract dilemma of what one ought to do in the face of suffering abstractly is to try and make sense of suffering abstractly. This is a form of mastery, and another way to derive pleasure from suffering. Missing out on the intellectual mastery and pleasure is what one hazards when allowing suffering to be particularized and specific. The loss of mastery and pleasure may be one of the motivations for resisting Laub. Yet, one cannot understand a particular and specific suffering without open and respectful listening, even if one rejects and wants to soften Laub's radical sort of strong autonomy.

References

1. Walzer M. Just and unjust wars: a moral argument with historical illustrations. New York: Basic Books; 1977.
2. Hobbes T; MacPherson CB, ed. Leviathan (1651). New York: Penguin Books; 1968.
3. Nietzsche F; Golffing F, trans. The Geneology of morals (German, 1887). Garden City, NY: Anchor; 1956.
4. Herman JL. Trauma and recovery: the aftermath of violence from domestic abuse to political terror. Basic Books; 1992.
5. Langer LL. Holocaust testimonies: the ruins of memory. New Haven, CT: Yale university; 1991.
6. Borowski T; Vedder B, trans. This way to the gas, ladies and gentlemen. In: This way to the gas, ladies and gentlemen (Polish, from 1948 story collection). New York: Penguin Books; 1967:29–49.
7. Laub D. Bearing witness or the vicissitudes of listening. In: Felman S, Laub D, eds. Testimony: crises of knowledge in literature, psychoanalysis, and history. New York: Routledge 1992:57–74.

Cross-Cultural Perspectives on Violence:

Is Violence Always Violence?

ARNOLD R. EISER, M.D.

UNDERSTANDING THE CULTURAL BASIS of violence requires a definition of violence. Defining violence depends on the social and cultural matrix that underlies the legitimization of an act of physical coercion within a society. Talcott Parsons observed that a society uses the threat of legitimate force to encourage conformity to regulations and as a symbol of security in the social system (1). Violence, on the other hand, undermines both the sense and reality of security. Grundy and Weinstein have added that violence damages the expected social structure by undermining the requisite sense of security (2). A government that enforces policies that are venal and unjustified betrays the society's trust and constitutes a particular type of violence.

Social theorists such as Rosenberg and Mercy (3) or Reid (4) identify four general categories of violence: 1) cultural, 2) structural, 3) economic, and 4) interactionist. Cultural causes include the normative values imbued in members of that society through socialization. In transcultural comparisons all four factors matter, but cultural valuation predominates. Structural factors influencing proclivity toward violence include family structures such as the absence of fathers. The socialization of males is particularly important to understanding the phenomenon of violence because of the marked male predominance in the commission of violent acts. Violence against females is particularly important both medically and sociologically. This paper will focus on an example of culturally sanctioned violence against women, female genital circumcision (mutilation), and critique its basis.

Kohlbergian Stages of Moral Development and Transcultural Comparisons of Violence

To compare violence in differing cultures it is helpful to use Kohlberg's theory of moral development as a heuristic tool (5). Kohlberg's thesis holds that there are six stages of moral development through which the individual passes progressively as he/she matures. Stage I is based on obedience to avoid negative reinforcement; Stage II consists of mutual cooperation motivated by personal self-interest; Stage III represents conformity to a code of conduct within one's social relationships; Stage IV comprises the social agreement which extends beyond one's personal relationships and involves rights and duties within a societal context; Stage V encompasses certain individual rights, such as life and liberty, which take precedence over majority opinion; Stage VI involves an objective criterion of universal applicability that takes precedence over social accords and a sense of validity of universal moral principles that prevail.

Cross-cultural studies using the Kohlbergian model have been summarized by Snarey (6). These studies reveal that, within given societies, attaining a given stage tends to correlate with the age at testing, i.e., older subjects were more likely to be in Stage IV or higher. Primitive societies such as Papuan New Guinea usually only reached Stage III whereas Eskimos did not reach beyond Stage II. Only middle-aged Israeli kibbutz members, Taiwanese graduate students, and middle-aged upper middle class people in India reached Stage V in appreciable percentages. In middle class Muslim families in Iran groups from age 9 to

---

From Elmhurst Hospital Center and the Mt. Sinai School of Medicine of the City University of New York. Address reprint requests to Arnold R. Eiser, M.D., Associate Clinical Professor of Medicine, 79-01 Broadway, Elmhurst, NY 11373.

adult scored at Stage IV, failing to show the usual age distribution. Scores among urban dwellers from different countries were more similar than their rural compatriots. Stage VI was not measurably present across a wide cross-section of cultures.

Some conclusions are evident from these studies of the Kohlbergian model. The Kantian nature of Stages V and VI accounts for their rarity and low distribution transculturally. The absence of Stage VI from every group studied raises questions of its validity as a category of moral developments for groups and individuals. Despite these limitations, I postulate that the framework of the theory will make cross-cultural comparisons of value judgments regarding violence more understandable. In developing a cross-cultural critique of female circumcision (mutilation), it is necessary, therefore, to ground it in an ethic that is not Neo-Kantian.

Female Genital Circumcision (Mutilation)

In Africa and the Middle East the practice of female genital circumcision is a widespread cultural practice. It is purported to regulate female sexuality and encourage chastity. Although the practice is found in predominantly Islamic regions, there is no Islamic doctrine supporting it. Non-Muslims in these regions also engage in the practice. Because of the recent substantial immigration to Western countries from countries where female circumcision is practiced and because this practice is extremely repugnant to Westerners, it has received much attention recently.

Stephen James argues that the practice is a violation of human rights and contrary to international law (7). Moreover female circumcision is a violation of the right to health because its mutilating character engenders illness in the form of clitoridectomy, infibulation, infection, psychic trauma, and complications in childbirth (8). Moreover because the procedure is generally performed on females before an age at which they can consent, it also has been viewed as child abuse. However, James goes on to claim that the core of international human rights is universal protection of individual autonomy. This is a Stage VI argument in Kohlberg's classification. As mentioned previously, no group studied in Western or other cultures scored at this level of moral development; hence, it constitutes an inadequate basis for a transcultural critique. In other words, basing an argument against female circumcision on Neo-Kantian principles is guaranteed to fall on "deaf ears" in a multicultural setting. The populations from localities where female circumcision is practiced have generally scored at the Stage IV level at which considerations of social accord and system maintenance prevail. For a people whose moral development conforms to this stage, the concern for conformity to tradition is the paramount concern. This is not to say that the practice is morally defensible. Rather, I suggest that it is possible to criticize this practice on grounds that are intelligible and persuasive to those who engage in it. To do so, one must appeal to transcultural moral standards that do not depend on Neo-Kantian standards of universality and autonomy.

I submit that health and well-being provide a transcultural standard. Having observed this *lingua franca* of health while practicing medicine for a decade and a half in the most international community of New York City, I have witnessed its validity. The yearning for health has a force that transcends particular cultural differences. Regardless of country of origin, patients desire an expeditious restoration of their health. An international moral standard prohibiting acts that cause illness does not require a Stage VI (Kantian) assumptions. Because illness, debility, and even death result from the practice of female genital circumcision (mutilation), it is morally unacceptable by this standard.

Although individual and community interests need to be balanced, a non-Western-centric morality can still reject such a practice. Those who defend the practice bear the burden of proof, because the harm of the practice is documented objectively, whereas no benefit has yet been shown. If benefit were demonstrated, such as a significantly lower incidence of AIDS in the populations that engaged in the practice, that benefit would have to be weighed against its health harms.

Recognizing health as a transcultural norm still leaves us with the task of adjudicating conflicting health claims. Concerns for the health of the person and the population must be reconciled. The conflict between community interests versus individual human rights also enters the consideration of this practice. For example, if certain females are particularly afflicted by this practice but the majority is not, then the conflict over individual rights is heightened. Western ethics places a priority on autonomy that is often missing in other cultural traditions, but adopting a thicker notion of the moral imperative for health raises the argument above the issue of personal autonomy and vouchsafes individual health from another perspective.

Health as a Transcultural Value

I have submitted that health is a transcultural moral standard. The universal longing for health and well-being offers two lessons for bioethical discourse. First, the striving for personal well-being takes priority over autonomy and control of decision-making. When push comes to shove, patients of diverse cultural backgrounds desire that the best health care decision be made in their case, not the most autonomous. This priority is reflected in studies of transcultural, nonmedical decision-making (9). Individual choice and informed consent can remain cherished principles but they should not be seen as preemptive.

The second lesson is that health care invariably concerns the community. It is not strictly a private, individual concern. Western minds may perceive this more readily when they contemplate far-off famines in the developing countries. It is no less true here in the United States. There is no better example than violence itself. Epidemics of violence involve the society as a whole and not merely the individual victims and perpetrators. Cultural norms determine the extent that a society tolerates and encourages organized and individual violence.

Conclusion

Politics, culture, and ethos determine which acts of physical coercion are prohibited within a specific society. Violence consists of physical coercion that remains outside of culturally sanctioned norms. In having a transcultural dialogue on morality, Kohlberg's stages of moral development help define the requirements of a transcultural moral standard. The practice of female genital circumcision (mutilation) powerfully illustrates the divergence of differing cultures. In the effort to critique this practice in a manner that "travels" outside Western presumptions, I have posited that health is a transcultural value that can be applied to both Western and non-Western systems. This claim is based on the experience of health care in a transcultural setting and is charged with a commitment to integrating prescriptive statements with descriptive ones. Health and well-being have ethical connotations that permit one to develop more meaningful transcultural moral statements than is possible with the Western concept of autonomy as the arbiter of moral truth.

Culture defines the format of daily life and life goals. It is a major determinant as well as deterrent of violent behavior. It defines what coercion is acceptable in life. Nevertheless, in our multicultural world we can yet reach out for transcultural norms and ideals that do not have to be reduced to Neo-Kantian values. Health and well-being is one such ideal but it must also be construed in a manner that balances communal, societal, and individual considerations. Female circumcision (mutilation) is an example of how health can serve as a transcultural moral norm to critique a morally objectionable practice in terms that can be understood and accepted in a variety of cultures.

References

1. Parsons T. Some reflections on the place of force in the social process. In: Eckstein H, ed. Internal war. New York: Free Press; 1964: 63.
2. Grundy KW, Weinstein MA. The ideologies of violence. Columbus: C. E. Merrill; 1974: 18–24.
3. Rosenberg ML, Mercy JA. Assaultive violence. In: Violence in America: a public health approach. New York: Oxford University Press; 1991: 25.
4. Reid ST. Crime and criminology. New York: Holt, Rhinehart and Winston; 1979; 173–224.
5. Kohlberg L. The philosophy of moral development. Moral stages and the idea of justice. San Francisco: Harper Row; 1981.
6. Snarey JR. Cross-cultural universality of social-moral development: a critical review of Kohlbergian research. Psychol Bull 1985; 97:202–232.
7. James SA. Reconciling human rights and cultural relativism: the case of female circumcision. Bioethics 1994; 8:1–26.
8. Toubia N. Female circumcision as a public health issue. N Engl J Med 1994; 331:712–716.
9. Haidt J, Koller SH, Dias MG. Affect, culture, and morality, or is it wrong to eat your dog? J Pers Soc Psych 1993; 65:613–628.

The Psychology of Violence

Robert Jay Lifton, M.D.

In talking about violence, we take on a subject that exceeds our capacity for anything like resolution. What we can and must do is address it in a constructive way and, I would add, bear witness to it. Our task is Sisyphean. If you remember, Sisyphus never could roll that boulder up the hill, yet he became a symbol—for Camus and others—of a certain kind of sustained dedication to a necessary task.

In psychological terms, being a witness means to take in and to open oneself up to an experience, to recreate it, and to build some kind of narrative about it. Being a witness requires being there, what existentialist Gabriel Marcel calls "presence." To attain this, the whole self must be present.

One of the difficulties with the subject of violence is that although one can talk about it, it is hard to discuss fully. One reason for this is that we tend to extract the death from violence, and put it aside. However, it is never put aside; it is always there. In thinking about bearing witness in a more systematic way, one comes to see that the narrative we build has something to do with justice. We are not just neutral bystanders when we bear witness; we are involved, we are part of the process. Holocaust survivors are often called collectors of justice because, as witnesses, they psychologically require the reinstatement of a moral universe. As physicians and medical professionals, when we experience violence we witness it in this way. As witnesses, we are or should be part of the struggle to reassert, or to assert for the first time, something approaching a moral.

My model for all work and for this kind of work is what I call advocacy and detachment. One's advocacy should be clear, strong, and displayed openly, but one's detachment has to be sufficient to allow for the application of professional knowledge in combination with advocacy.

I will discuss the different levels of violence, keeping in mind the nonneutrality of witnessing. In that spirit, I take you on a bit of a journey through violence, a journey that has three primary parts: the first part contains some psychological concepts about violence, the second part contains some thoughts about individual psychological experiences of violence, and the third part includes the more collective forms of violence or mass violence. As a kind of coda to the discussion, we need a concept of species self and of species mentality that could lead to alternatives to violence. This immersion in violence is not an effort of despair; it is an act of hope. In such an immersion one is opening oneself up to violent events for the purpose of witnessing them and seeking alternatives to them.

The Psychological Frame for Violence

The Common Quality in Violence

Years ago, during the height of the Vietnam War, my wife and I went to Vietnam. We had been in the Far East and decided that we would visit that country insofar as one could still do that, given the war. My wife went toward the front because she was studying the whole question of civilian casualties, including casualties among children. When she was walking toward the front, she came upon a black GI from New York. They chatted for a while and she asked him, "Is it very dangerous where you are?" He thought for a moment and he said, "Well, I'll tell you, it's a little more dangerous than 110th Street but a little less than 125th Street."

This story has many meanings, of course. First, it suggests that there is a common quality

---

Distinguished Professor of Psychiatry and Psychology, John Jay College and The Graduate School, City University of New York.

in all forms of violence. It also describes certain neighborhoods as war zones, suggesting that there are zones or locales of violence that do not depend primarily on individual psychology but rather are enclaves in which violence is a form of discourse. It also suggests one of the principles that I want to express in this paper: that as different as the levels of violence are, consistent human threads run through them all, each presenting different problems for the world.

The Life (Vitality) and Death (Immortality) Continuum

Another principle of violence has to do with the approach one takes. For a psychological approach to violence, a life/death model or paradigm is needed. One cannot eliminate death from psychology and still talk about violence. Much violence can be seen as a reaction to a perceived threat of death to the person, to the group, to the self, to the psyche. Such a threat or idea can be a perception under different conditions, but the death imagery is crucial. In my work I move away from the classical Freudian model or paradigm of instinct and defense to one of the continuity of human life or the symbolization of life and death.

Much of violence is a quest for vitality. That is what makes this issue so difficult. There is often a struggle to overcome one's inundation with the imagery of death. For instance, studies of young criminals show that they are somewhat depressed, numb, and withdrawn; then they commit a violent act and for a time experience considerable vitality. After a while the vitality ebbs, and they sink back into the slightly depressive or numb stage. Thus, violence can become, under some conditions, a quest for vitality.

In this model of death and the continuity of life and the symbolization of life and death, there is also a larger psychological domain, that is, the proximate idea of nitty-gritty psychological experience. Also there is the area of larger human continuity, what I call the symbolization of immortality. It is central to all of our lives, but insufficiently considered. The issue of immortality has been left too much to the philosophers and theologians who have done much with it, whereas we in psychology have done too little. Immortality has to do with our need, our psychological requirement, for a larger sense of continuity, a sense of living on in our children or their children, or in our works or what we create in the world or in our effects on other human beings. It may even be realized through some spiritual principle of a major religion, or through eternal nature, or through a psychic state, an experience of transcendence, a state so intense that within it time and death disappear. These are all modes of larger human continuity or the symbolization of immortality. They have a bearing on violence because these expressions of continuity can be sought through violence.

The Experience of the Violent Self

Another crucial principle in looking at the psychology of violence is the experience of a self, the experience of the person perpetrating the violence. We shy away from this because it is unpleasant. I felt such unpleasantness when I was interviewing Nazi doctors. One does not like to talk to people who have committed ugly acts of violence. But one has to, and one has to follow up on those talks and put together some kind of picture.

Ideology

A third very important principle is that of ideology. Ideologies of violence are crucial to violence in any degree. Americans tend to downplay ideology; we say we are not ideological people, but there is always an ideology of violence, what I call a violence construct, where violence is being committed. Those who commit suicide have, in addition to despair, a suicide construct, in which from early in life they say, "I am someone who can kill myself." It is part of an identity, a sense of self. The same is true of those who commit violence—violence can become part of an identity, part of a sense of self, and in that way one can speak of a violence construct.

Passionate and Numbed Violence

In reviewing the principles, one might think that there should be a distinction between passionate and numbed violence. For instance, we think of family violence as passionate—it is people who are close to each other who can become enraged at each other—whereas we think of gang violence as numbed, as akin to the out-of-town killer who is a stranger. A killer from out of town is hired as a hit man because he will not know the family or the children who will suffer, or any other details about the person he is employed to kill. The hit man is an example of numbed violence, as is the technological violence that has been pursued in all recent wars. Atomic violence is numbed violence taken to its extreme.

Immediate violence is both numbed and passionate. In family violence there is, of course, a passion, but at the moment of killing, the other person is depersonalized and dehumanized, and the killing is done in a moment of coldness. Without that dehumanization the killing could not take place. In various other events in which the violence seems numbed, there is passion behind it that creates the act of violence. There is always that combination.

Vietnam Veterans

I gained what I thought was some understanding of violence and the psychology of the individual violent self when I worked with Vietnam veterans and antiwar veterans in the early 1970s. We formed rap groups—they weren't called therapy groups then. They were exciting, interesting, egalitarian experiences with give-and-take from all sides. When we talked about violence and explored it, the veterans raised three themes.

One theme was what I call the habit of violence. There can be certain situations in which violence is the normative form of discourse. It is the way one communicates, the normal mode of behavior in that particular environment, an environment that exists in war in general and especially in a counterinsurgency war like the Vietnam War. This war was perceived to be an absurd war by many of those who were fighting it. In such a situation one reacts to any kind of surprise or threat with immediate violence, as indeed often happened there.

A second theme was that violence can be a response to betrayal, or, in the veterans' more colorful language, being "fucked over" by their own government. Betrayal involves having been drawn to something or someone one loved or supported and then being tossed aside, deceived, badly used, or in some way exploited. Many people who fought in Vietnam felt that way about their government.

A third theme was the experience of being or feeling thwarted. It was interesting to watch these young men change as we worked with them because at times various events would happen that would thwart them. One young man described an experience of giving a talk before a group of elderly war veterans and telling about the horrors of Vietnam and the absurdity of the war. One elderly veteran stood up and said, "Look, what you said is all true. But I don't care about gooks, and I'm a white man, so it doesn't mean a thing to me." The speaker was a young man who was struggling hard to extricate himself from similar feelings and who had changed a great deal. He described becoming so enraged at the time of being thwarted in his effort to change himself that he felt like blowing up the whole building. To address this sense of deep frustration, he discussed the experience in our group.

The Image and the Act

This example raises the issue of distinguishing the imaging of violence from the act itself. This does not mean that the imaging of violence can't be dangerous—it can be, but one must make a distinction in the sequence. The imaging of violence—for example, thinking "I could kill him"—is something most of us have experienced. But in carrying through with this action, a threshold is crossed, and it is not always understood how that happens, how someone comes to cross such a threshold. At an individual level it may relate to what the self can tolerate.

Attractions to Violence

An important study that describes some unpleasant truths about individual acts of violence is called *Seductions of Crime*, by a sociologist in California, Jack Katz (1). In it he discusses some of the psychological attractions to violence. He describes moral self-transcendence, the feeling that one is not ordinary, or the feeling that one is on an extraordinary and exciting mission and so not bound by mundane authorities and rules of conduct. There is a description in his book of a gang in Glasgow, whose leader says, "See, you feel it in your belly as you're going into battle. It is like a feeling you have when the Rangers attack the Celtic goal. Your heart's racing, you feel sick. It's better'n sex." Violence can create that excitement in some people, but often the seeds for that kind of violence are sown in a communal process, a social setting. Another phenomenon, occurring mostly in men, is creating oneself as the so-called hard man, the "barrio warrior," the "bad ass," the "wise guy." These are the tough killers who will kill and become heroes to those around them because they show no remorse and virtually no limits on what they are willing to do to assert themselves through violence. They take on a certain leadership in groups and also develop disciples and followers.

Righteousness

In carrying through with a violent action there can also be a sense of righteousness, righteously enraged slaughter. Very quickly one builds the construct to inwardly justify the act one is committing. "He deserves it! He was a lousy husband." "She deserves it, she wasn't a good mother." This concept is very important in understanding the individual psychology of violence. We interviewed people at our center, the John Jay College Center on Violence and Human Survival. One man told of how he was brought up in the ghetto always thinking of himself as a punk, a zero, a nothing, until he acquired a gun. This is not an unusual story. He said, "Once I got a gun at the age of 16, I figured, if I have a gun, you fear getting shot. Rather than a zero I am somebody if I have a gun." Similarly with drugs, one imagines oneself to have control over the substance that is in fact doing the controlling, but there is the sense of moving toward the ideal of the "bad ass" as an identity. In some subcultures this is one of the few identities available.

Violence as Pride

Violence can also be a matter of pride for some people, an individual psychological experience. One man interviewed by Jack Katz describes making a call to his son. He was someone who had been a gunman and involved in various kinds of violent activities. He warned his son over the phone saying, "Look, I'm going to beat you up when I get home." His son asked him, "Why are you going to beat me up? I haven't done anything." The man explained, "I'll do it because I feel like it." The man told Katz, "He will know that it won't be no ordinary beating, no ordinary hitting on him. This is part of him growing up and a part of me coming out that I want him to inherit from me." (1) He wanted to transmit to his son the power that a violent person can feel and to maintain what I call his symbolic immortality through that transmission of violence. It is a tenuous and undesirable psychological adjustment, needless to say, but one should look at it as what results when violence becomes the norm and is rewarded in various social ways.

Group Ethos

At a collective level there can be a shared group ethos of crossing over. I heard a very powerful description of what happened with the men at the My Lai massacre in Vietnam. Apparently they talked at great length about what they would do to the gooks, how they could do it. Through talking about it, they reached a point where encouraged by each other and by the group ethos, they essentially did what they had imagined. What happened at My Lai applies not only to My Lai. It is part of what I came to describe as an atrocity-producing situation, in which individual psychology and the collective psychological field that the individual moves in are intermixed. I would define an atrocity-producing situation as one in which the average person can enter into the "psychology of slaughter" given the prevailing external conditions (2). At My Lai, not only had the men been in Vietnam for a while but they also underwent a set of circumstances which could be considered atrocity-producing situations. They began to lose men in a horrible way to land mines. There is some question as to who set those land mines, but horrible deaths occurred because of them, including the death of a sergeant who was a father figure. The men were frustrated, angry, and terrified, and the night before My Lai there was a combination religious ceremony for the dead sergeant and a kind of rally. People differ on exactly what orders were given, but in the briefing the idea that one kills everything and everyone was conveyed. As this informant described to me, at My Lai they felt as though they were in combat. They killed little children, old men, old women—five hundred people in a matter of a few hours. One could surmise from the description of the event that the soldiers were under the illusion that they had finally gotten the enemy to stand up and fight and were in combat for that moment.

I was very intent on understanding how one young man at My Lai developed restraint not to fire. He not only didn't fire, but he pointed his gun very visibly to the ground so the group knew he was not firing. (A few others who did not fire at anyone pretended to shoot so as not to be ostracized by the group.) In my interviews with him, three factors that influenced him could be discerned. First, he had had a Catholic upbringing, and although he had broken with the church he still remembered some of the teachings of the priests and the nuns which placed restraints on violence and on what one could do to other people. Second, he was a loner. He had grown up by a seaside and had gone boating by himself. He did not really like to be with other people, and therefore was not easily influenced by other people. He did not get into the group ethos that scapegoated the so-called gooks and victimized the Vietnamese. Third, and perhaps most interesting, he had joined the military after dropping out of college. In college he had been doing poorly, he was feeling confused, he was drinking excessively, and he did not know what he wanted to do with his life. In the military, he found himself. He loved the military. He excelled at the training exercises and quickly earned some of his stripes. He was then honored everywhere he went and he thought he had finally found his place. He went to Vietnam, and then at My Lai he was horrified at what happened. He thought it unworthy of his idealized sense of the military; this served as a major source of his restraint.

Violence in America

My Lai was an atrocity-producing situation and we create My Lai-like situations in our inner cities and elsewhere. One must think psychohistorically about these situations to discover some of the sources of American violence. Richard Slotkin wrote a whole series of books about regeneration through violence (3). He pointed out that the real creators of this country were not the gentlemen from Boston and Philadelphia, but rather the tough characters, the rogue adventurers. They were land-boomers, Indian fighters, traders, missionaries, explorers, and hunters, who killed and were killed until they had destroyed many among the native population and mastered the wilderness. Often they would be initiated by Puritan pilgrim groups at a time when new generations were relaxing their religious devotion. Thus, the violence was a form of regeneration, or, as I mentioned earlier, a form of achieving vitality or of reasserting larger symbolic immortality in that regenerative fashion. Much of America is built on that history. We are still stuck with some of the convictions of that early time, particularly our obsession with guns, which creates violence itself, not just violent impulses.

The gun issue in American society is very powerful, and very important, in terms of the dangers presented by allowing so many guns of such lethal quality into our society and in terms of what it symbolizes about our capacity, or lack of capacity, to cope with violence. I had many discussions with the very talented American historian Richard Hofstadter who died 25 years ago. He said what had discouraged him most was the American inability to cope with guns, to take appropriate actions about guns. I think he was expressing the tremendous significance of guns for American violence.

Writers like William Julius Wilson, Kenneth Clarke, and Cornell West talk about ghetto areas or truly disadvantaged areas as being in some ways severed from the rest of society and victimized. Cornell West and others described the long historical sequence of slavery, postslavery, and white victimization that led to a breakdown of black culture. There are also violent subcultures among poor whites. The novels of Russell Banks, particularly *Continental Drift* (4) and *Affliction* (5), are about violence as a form of discourse among the white poor who have lost their moorings. Life has lost meaning and, not knowing where to turn, they seek meaning in violence.

Part of the problem in trying to stem violence is the transmission of macho male violence through the generations from father to son, as a way of trying to hold oneself together. In my recent book, *The Protean Self* (6), I discuss an emerging sense of the self that has elements of flexibility and many sidedness. It is in all of us to some degree, although it needs some steadying with elements of stability or form, and most of us struggle to make the protean self work in one way or another. However, when the protean self is added to an already fragmented self, as is typically found in these subcultures of poor whites or ghetto blacks, then the fragmentation goes further and there is further inner chaos. Resorting to violence may be an effort to re-create some degree of form as in violent gangs that have their own subsocieties, creating forms and rules.

One of the sad things that we found in our interviews with some of the people from the ghetto was the lost talent. One man, who had been in prison for armed robbery, was sensitive, intellectually active, and a reader of books. He probed the question about justice for all people with great sensitivity but he could not hold onto a sense of self that incorporated his sensitivity. He is the man I mentioned earlier who became a somebody rather than a zero when he acquired a gun. The identity of the "bad ass" or the macho person in the ghetto was the only one that he could make work. It became heartbreaking to watch that process, all the more so because the other side almost took shape within him.

Dissociation

All of the kinds of violence require psychological forms of dissociation. Dissociation is a term originally used by Pierre Janet, a French psychiatrist and contemporary of Freud. Janet found there is a tendency to dissociate in the face of stress. Dissociation means the separation of part of the self from the rest of the self, so that one part of the self can seem to operate relatively autonomously. The psychoanalytic term is splitting. In my work with Nazi doctors, I use the term doubling, because I found the Nazi doctors formed a second functional self to perform their violent and their murderous activities in the death camps. They were ordinary men, husbands and fathers. They would return home a few days a month to visit their families in Germany. But once back in

Auschwitz, they would do their dirty work and their killing. They had formed an Auschwitz self, a double of their ordinary, relatively humane, prior self.

This doubling is not limited to Nazi doctors. One finds it in criminals and in people who generally live an ordinary life but at times engage in killing, as do some members of the Mafia or terrorist groups. Doubling is a form of dissociation as is numbing, a diminished capacity or inclination to feel. In large-scale killing, there is doubling in those who are most responsible for the actual killing, and numbing and other forms of dissociation in those who are involved but less responsible.

Violence in our Contemporary World

A common thread runs through the violence in the former Yugoslavia or in Rwanda, which are examples of genocidal violence, or at smaller expressions of violence, as in Waco, Texas, or the cult in Japan that released poisonous gas into the subway. They are totalistic societies or groups and their goal is some absolute. In the case of Serbian violence the ethnic cleansing is very close to the Nazi biomedical vision, and represents an absolute. There is a vision of purification in the most extreme forms of nationalism and there is something like an end-time image. With the Branch Davidians and the Japanese Aum Shinrikyo cult, the end-time imagery is overt. The end of the world is anticipated. Indeed Koresh and his Branch Davidians sought the end. There is a worldwide epidemic of fundamentalism in large and small groups and much of it is characterized by end-time ideology. Some of these groups become violent, and there is a potential for a type of holy war violence in these groups; they need to be seen in that way. In one sense fundamentalism is a reaction against proteanism. Proteanism is bequeathed to us by historical changes and historical confusion but it is very difficult for many people to live with so much uncertainty. Some seek a plunge into an envisioned past, a perfect harmony that never was. They worship the sacred doctrine of fundamentalism and often add an end-time anticipation from the Book of Revelation and may even seek the end.

Genocidal Mentality

I wrote a book with Erik Markusen called *Genocidal Mentality* (7). In it I developed some ideas about the Nazi sequence. I applied them in some parallel way to the nuclear sequence. The genocidal sequence, which applies in Bosnia and Rwanda, involves the perception of a great historical trauma that gives people a sense of despair, as in the case of the German defeat in World War I. After the trauma there is an attempt to overcome and revitalize oneself through some powerful ideology. The ideology can become genocidal, as in the case of the Nazi biomedical vision, or the embrace of nuclearism. Interestingly, in this process, intellectuals and professionals become central to large-scale violence. They provide the ideologies, the technologies, and most of the organizing. The principle of ethnic cleansing came from the Serbian Academy of Sciences.

The American sense of trauma after the bombs were dropped in Hiroshima and Nagasaki created what was called a cultural crisis in the United States. That experience is examined in a wonderful book by Paul Boyer entitled, *By the Bombs' Early Light* (8). In the case of a nuclear threat, the trauma is by no means gone, merely dispersed. One of the difficulties now is that we are no longer facing nuclear genocide vis-á-vis the Soviet Union. We have lost our nuclear fear, to some extent. In our interviews we still encounter some of it, but it is a shadow and it merges with environmental fear. We do not have an appropriate kind of fear that would help us to do what we should without nuclear weapons.

The developing species mentality or the species self is an alternative to violence. One of the hopeful aspects of the protean self is that it can open out toward a species self. What I mean by species self is not some idealistic floating sense of being among some universal principles, but rather a sense that one is a member of humankind. One is a human being. It does not mean we cease being what we are. If I have any such species self in me I am still an American, a Jew, a teacher, a psychiatrist, a tennis player, etc. This is a very practical, concrete matter because my sense of being a human being will affect how I look at the world and the positions I take about world events. In our better moments, for instance, as we watched those extraordinary revolutions in 1989 and 1990, we were not just watching the Chinese marching for democracy or South Africans marching for democracy or the Czechs in some remarkable way achieving their velvet revolution. It was fellow human beings we saw liberating themselves and we felt vicariously further liberated ourselves. The same process occurs when one sees a baby starving in Africa. It is not an African, it is not a black, it is a baby. We feel certain emotions about babies; they are the kind of helpless, very young human being that we are supposed to protect and nurture, and that is an expression of species mentality. Although the mass media do all kinds of dreadful things, one of their more hopeful potentialities is that they bring before us more possibilities for a species mentality.

Doctors and Violence

When I was very active in the doctors antinuclear movement, I was both studying Nazi doctors and working closely with physicians for social responsibility and with international physicians for the prevention of nuclear war. This was a very good juxtaposition of two different directions doctors could go in and I could not help but be struck by the contrast. We made a trip to the then Soviet Union with the doctors' movement. During the height of the Cold War, a very sick man was in a large Soviet cardiology center. It so happened that the heads of both of our delegations were cardiologists, and cardiologists have a heart. Each one examined this very sick man in turn, putting a stethoscope on his chest. Suddenly as I watched, "Americans" and "Soviets" disappeared; even the nuclear weapons issued disappeared. These were just two healers examining a very sick man in turn and trying to think of some way that they could contribute to prolonging his life. This scene makes one think about the species mentality component of the medical profession. We sometimes lose it with all the ambitions we have or with our corruptibility. But in law the species principle is justice. In science and academic life it is truth. And, of course, in medicine it is healing. So, I would say, we are as physicians and medical professionals in general, quite capable of the rediscovery of that commitment to healing as a means of witnessing and combating the violence around us and in us.

References

1. Katz J. Seductions of crime: moral and sensual attractions in doing evil. New York: Basic Books; 1988.
2. Gault BW. Some remarks on slaughter. Am J Psychiatry. 1971; 128:450–451.
3. Slotkin R. Regeneration through violence: the mythology of the American frontier, 1600–1860. Wesleyan University Press; 1973.
4. Banks R. Continental drift. New York: Harper and Row; 1985.
5. Banks R. Affliction. New York: Harper and Row; 1989.
6. Lifton RJ. The protean self: human resilience in an age of fragmentation. New York: Basic Books; 1993.
7. Lifton RJ, Markusen E. Genocidal mentality: Nazi holocaust and nuclear threat. New York: Basic Books; 1990.
8. Boyer P. By the bombs early light: American thought and culture at the dawn of the atomic age. Pantheon; 1985.

Violence against Elderly People

Barbara E. Cammer Paris, M.D.

Abstract

Every year, more than one million older Americans are injured physically, debilitated psychologically, or exploited financially by a family member. Elderly men and women, who are from all socioeconomic, ethnic, and religious backgrounds and who have varying functional abilities are vulnerable to abuse and neglect. Health care professionals need to develop the skills to detect both patients at risk and those actively being abused.

Barriers to detection of abuse and neglect are complex and include victims' resistance to disclosure due to their own ageist attitudes or feelings of shame and guilt; isolation of victims by their abusers; and the health care professional's discomfort with the situation and hesitancy to intervene, often as a result of inadequate knowledge and training in this area.

Researchers have identified specific risk factors for abuse and neglect by family members, including psychopathology among family members, a family history of transgenerational violence, the elder's dependency, the elders and the caregivers' isolation, the caregivers' stress, and living arrangements. Thorough assessment of patients at risk by a multidisciplinary team including a physician, a nurse, and a social worker and the team's subsequent development of individualized intervention strategies can have a positive impact on this devastating problem.

An 80-year-old woman was examined regularly by a geriatrician for congestive heart failure. Despite the physician's belief that the patient's condition should have been well controlled by the medication prescribed, the patient was repeatedly hospitalized for heart failure. The physician was puzzled about the reasons for the hospitalizations. While the patient was hospitalized, the physician was able to visit the patient daily and establish trust and rapport. Ultimately, the patient confided in the physician that she was not taking her medications at home because her 47-year-old daughter, who was living with her and depended on her financially, was using most of the patient's social security check to purchase illicit drugs. The patient was unable to purchase the prescribed medication. Feeling shame, embarrassment, humiliation, and blameworthy, and fearing retaliation from her addicted daughter, she never had confessed this information to anyone.

Extent of the Problem

A 1986 report by the Subcommittee on Aging of the U.S. House of Representatives estimated that each year more than one million older Americans are injured physically, debilitated psychologically, or exploited financially by a family member (1). Applying this estimated rate of abuse to the elderly population in New York City suggests that approximately 40,000 seniors in the city are abused each year. By the year 2030, there will be about 70 million older Americans, more than twice the number in 1990, and a proportionate increase in the number of older victims of abuse can be expected (2).

Like the patient presented above many elderly individuals are reluctant to discuss their problems. Moreover, many of these individuals sense

---

Assistant Professor of Geriatrics and Medicine, Medical Director, Coffey Geriatrics Clinic, Mount Sinai School of Medicine, New York, City University of New York.

that many professionals are uncomfortable in addressing questions of neglect and violence for a variety of reasons, including the absence of available resources to help solve the problem.

The crucial initial step in the process of overcoming the problem of abuse and neglect of the elderly is recognition and acknowledgment of the problem. The next step should be developing the skills to detect active abuse and patients at risk. Professionals can play a critical role in detecting, assessing, and intervening in abuse and neglect of the elderly especially when it is most likely to occur at the hand of a family member and go unnoticed.

Types of Abuse and Interventions

Abuse can occur either as a continuing pattern throughout life or as a late-onset event. The literature offers no universal definition of the term “elder abuse.” Although no agreement has been reached on what constitutes abuse or what ages are encompassed, we can categorize elder abuse as being psychological, financial, and physical, intentional or unintentional (3). Often, several types of mistreatment occur simultaneously; and, over time, the abuse is likely to escalate in both frequency and severity.

Categorization of the type of mistreatment is helpful in planning intervention. For example, neglect could be unintentional when an illiterate spouse caring for a frail, elderly, medically complex patient cannot read the label on the medication bottle. Incorrect dispensing of the medication may lead to an adverse side effect for the patient. In this instance, the intervention might be educating the caregiver on when to administer the medicine and in what quantity. In other situations, there is deliberate intent to harm and to inflict physical pain or emotional suffering and/or to take advantage of the other person to fulfill one’s own needs.

Bars to Detection

The Victim

The victim’s feeling of embarrassment or guilt about the abuse is often a significant barrier to detection. The victim may also worry about retaliation or some form of punishment by the abuser as a consequence of exposure. Moreover, the victim may fear institutionalization, in the form of nursing home placement, as the only alternative solution. Although some victims may be willing to discuss their abusive situation with a health care provider, victims often minimize the seriousness of the abuse and display tremendous loyalty to their abuser. Victims usually are not willing to take legal action against a family member. Many victims prefer tolerating the abuse to losing close personal relationships in their lives. For example, in the case I have presented, the mother was actually more concerned about obtaining drug rehabilitation for her daughter than about her own potentially life-threatening situation.

Isolation of the victim is another serious barrier to detection of abuse and neglect. Elderly victims may become secluded as abuse escalates, feeling uncomfortable with having visitors in their homes or in communicating with others about their problems. Abusers also often isolate elderly victims by denying them access to friends and the telephone and escorting the victim when outside contacts are made to ensure secrecy. Some victims are functionally impaired and physically dependent on the abuser for access outside the home. A visit to the hospital or doctor may be the only contact the victim has with the outside world for several months or even for years.

The victim’s own ageist attitudes can also result in failure to recognize an abusive situation. A victim may have the attitude that old people are a burden to their family and therefore deserve to be neglected or even mistreated. For example, a physician learns that an unkempt, disabled, malnourished patient is being cared for by his son who has been too busy to visit or help for the past two months. During this interval the patient has been trying to manage without any assistance, unable or unwilling to question his son’s actions or to ask others for help.

Health Care Providers

Lack of training of professionals in appropriate methods of interviewing patients and family members will result in significant underdetection of abuse and neglect. Most health care providers have not had structured training in screening for abuse or in methods of conducting interviews with potential victims and abusers. Health professionals feel tremendous discomfort in approaching the topic with their patients.

A final important barrier to detection may be the health care provider’s discomfort with and resistance to recognizing a potentially abusive situation. For example, a physician may respond to an 80-year-old patient complaining about his grandson stealing his check by tacitly denying the validity of the patient's accusation and suggesting instead that the patient may have misplaced the check. In some circumstances, doctors feel that it is not within their domain to intervene in a domestic situation, especially if the victim is not mentally impaired. They may fear involvement with litigation; they may also hesitate to pursue issues without knowledge of potential solutions.

Risk Factors

Researchers have identified specific risk factors for abuse and neglect by family members (4). These include:

- psychopathology in family members
- family history of transgenerational violence
- the dependency of the victim
- isolation of the victim or of the caregiver
- the caregiver's stress
- living arrangements

Awareness of these factors can assist health care providers in identifying patients who would potentially benefit from a comprehensive assessment (5).

Medical Assessment of Victim and Abuser

Medical assessment of the victim should include:

- a comprehensive history (patterns of obtaining medical care, consistency of history with medical observation)
- a complete physical examination including neurologic examination, gait observation and inspection of the genitalia, a cognitive assessment including the patient's ability to comprehend the risks and consequences of their actions
- a mood assessment for signs of depression or withdrawal, or fear without any clear signs of depression, or of cause and effect.

The next step in the process is written documentation of the assessment including an accurate account of the patient's history and physical findings and notation of any inconsistencies in the data collected, such as multiple vague differing accounts of events or injuries that are inconsistent with the history. The physician may also call on the expertise of a nurse and a social worker for comprehensive assessment.

In some situations it may also be feasible and important to do a formal assessment of the potential abuser. This assessment would include an evaluation of living arrangements, finances, and psychosocial stresses. Relevant information about the potential abuser might refer to drug or alcohol dependency and mental illness or disability. Financial dependence of the abuser on the victim for food and shelter may also affect their relationship adversely. On the other hand, the dependence of the victim on the abuser may result in overwhelming caregiver stress. The abuser should be interviewed in a nonjudgmental and timely fashion to avoid the abuser's collusion with the victim in coordinating responses to specific questions.

Intervention Strategies

Once the assessment process has been initiated, one can begin to work on intervention strategies for coping with elder abuse or neglect. Intervention, like detection and assessment, is often a multistep complex process which can be enhanced by a multidisciplinary team (6).

Options for action should be based on information obtained during the assessment. In instances of unintentional neglect, for example, physicians can play a key role in educating the caregivers about the needs of the patient and the nature of the illness. Detecting caregiver stress and suggesting ways to alleviate it can often avert a crisis. Aging individuals can be counseled to plan for future incapacity and dependence before they become vulnerable to abuse.

In cases of intentional abuse, it is important to ascertain how the patient feels about the abuse and if any plans to cope with the situation have been formulated or implemented. Additionally, one needs to determine how long the abuse has been going on, whether it is escalating in intensity or frequency, and whether the victim is aware of any factors that precipitate the abuse. This information will help guide the planning of intervention strategies. For example, if the patient is in a life-threatening situation, alternative living arrangements for the victim or the abuser need to be explored. If the abuse is ongoing and chronic, initial interventions may include education of the victim on emergency phone numbers for police, shelter and abuse hotlines, and the establishment of safety plans.

It is important for a health care provider to remain accessible to the victim who may initially reject offers of assistance; victims frequently fluctuate between the stages of denial and acknowledgment of the abusive situation. Access can be maintained via telephone monitoring or through more frequent visits to the doctor.

Summary

Detection and assessment of elder mistreatment can be tedious. It is highly unlikely that many elderly individuals will report acts of aggression against them by members of their families or by other caretakers on whom they rely for their basic needs. Therefore, it is crucial that health care professionals ask the right questions and perhaps read between the lines to single out those elders who are at risk of abuse, or who may be current victims of abuse.

References

1. U.S. Congress, House Select Committee on Aging. Elder abuse: what can be done? Washington, D.C.: Government Printing Office; 1991.
2. The Program Resources Department, American Association of Retired Persons (AARP) and the Administration on Aging (AOA), U.S. Department of Health and Human Services. A profile of older Americans. 1993.
3. Lachs MS, Pillemer K. Abuse and neglect of elderly persons. N Engl J Med. 1995;332:437–443.
4. Breckman RS, Adelman RD. Strategies for helping victims of elder mistreatment. Newbury Park: Sage Publications; 1989:35–36.
5. American Medical Association. Diagnostic and treatment guidelines on elder abuse and neglect. Chicago, IL: American Medical Association; October 1992.
6. Matlaw JR, Spence DW. The hospital elder assessment team: a protocol for suspected cases of elder abuse and neglect. Elder Abuse Neglect 1994;6:23–37.

Violence against Infants

Katherine Teets Grimm, M.D.

Anyone who listens to the news or reads the papers regularly knows that child abuse is a serious problem. The media document only a fraction of the worst cases. Some statistics from the Fatality Review Committee, which reviews all suspicious deaths of children younger than 5 years old in New York City, convey the magnitude of the problem. In 1993, nearly 52,000 reports of child abuse were made within the five boroughs of New York City. Child protection laws cover children younger than 18 years of age plus the small number of older adolescents and young adults older than 18 who are in residential treatment programs. Of the children younger than the age of 5 who died, 93 of the deaths were suspicious. Of the 93 deaths, 37 occurred in families already known to the child protection system.

The Fatality Review Committee looked carefully at those 37 deaths and made some interesting observations. Almost 50% of the fatalities resulting from child abuse were the result of blunt trauma. A large number of these fatalities occurred in children younger than one year of age. The average age of the mother of a child who died was just over 28 and she had 3.9 children. This raises the issue of teenage pregnancy and subsequent violence. Those of us who deal with the problem of teenage pregnancy have observed that the family rallies around and supports the mother with her first child. By the time the third or fourth child arrives, patience and support have evaporated. These women tend to be isolated.

A frequent factor in child deaths secondary to abuse is a history of substance abuse; two thirds of the children who died were in families whose members had this history. Substance abuse affects frustration threshold level. Some substances such as cocaine, PCP, crack, and alcohol are more likely to result in violent behavior and thus are more likely to produce child abuse

Twenty-five percent of the children who suffered fatal injuries in 1993 had positive toxicology screen tests at birth, indicating that the mother was using drugs during her pregnancy. This has become a politically charged issue. If a child has a positive toxicology screen and the child's mother claims she only took the drug once and there is no evidence of other problems, the child welfare system may not act. Other risk factors need to be recognized before they intervene.

The children who become fatalities are often in some way already damaged either by physical or developmental disabilities. Sometimes children become targets in a family because they are different. More than 40% of the fatalities were in homes where domestic violence took place. There is a strong link between domestic violence and child abuse. It is the considered opinion of most who work in this area that if a woman comes into the emergency room as a victim of domestic violence, the whole family system needs to be examined. In one third of the fatalities, the families of these children had been homeless within the past 3 years. One third had adolescent mothers and 20% had adults involved in their care who suffered from mental illness. Two key features stand out in the homicide cases: domestic violence and mental illness. I suspect that drug abuse was also a key feature.

Most children who are victims of abuse do not die from the abuse. In my work with children and adults, I can attest to the significant psychological impact that results from physical, sexual, and other kinds of abuse. The psychological scars that result from sexual abuse, something most children do survive, are especially profound. They are

---

Associate Professor of Pediatrics, Director, Pediatric Emergency Services, Mount Sinai School of Medicine, New York, City University of New York.

often the precursor to social problems, including early adolescent pregnancy. In our teen pregnancy clinic, the social worker has observed that many of the 13-year-olds who are pregnant were the victims of past abuse. Abuse plays havoc with one's sense of self. Victims of it often feel inferior and have self-destructive tendencies. These tendencies can turn inward, resulting in behaviors like drug abuse, suicide, and I would argue, teen pregnancy. In males the response is typically turned outward. It is not unusual for male victims to become involved in cycles of violence and abuse as social outcasts and criminals.

The cases I present involve deaths and unequivocal violence. A 5-day-old girl was brought to the emergency room dead on arrival. She had been born at The Mount Sinai Hospital to a 16-year-old mother and was the product of a full-term pregnancy. This mother remained engaged in a tumultuous relationship with her boyfriend, the child's father. On the day the child died, the mother had gone to the supermarket, leaving her child with the father. When she returned, she found the child dead in her infant seat. The boyfriend claimed he had been in the other room and the baby had cried for a bit but then stopped, and he assumed she had gone to sleep. The Medical Examiner was concerned because sudden infant death does not usually occur this early in life. The ultimate conclusion, although it was hard to prove, was that the child was smothered.

If there is someone in the household who is on drugs or mentally ill, a baby's crying can trigger violent behavior. In these situations we often see cases of smothering or shaken baby syndrome. When a baby is shaken violently, bleeding in the brain may ensue. The child either dies or suffers permanent neurologic damage.

The second case is similar to the first, only the child died at an older age. The mother of this child was also a teenager, 18 years old. We had observed her during the pregnancy in the clinic and had found her to be contentious. She too was involved in a tumultuous relationship with a man. Our social worker read about the fate of this child in the newspaper. At the age of 7 months, he was the victim of blunt trauma. He was beaten to death with a rattle. I suspect this was precipitated by crying.

The third case is illustrative of a pervasive problem in all communities, crossing all socioeconomic groups. A 15-month-old girl was brought to the emergency room dead on arrival. This child was too old for sudden infant death. The physicians in the emergency room were doing a routine resuscitation and part of the routine is to try to catheterize the patient to obtain urine. When the physician looked at the genitalia, he discovered huge rectal and vaginal openings. It was clear that this child had been the victim of ongoing sexual abuse. We could not say whether the death was related to this sexual abuse or not. If it were related to this abuse, it would have been death by suffocation. As I indicated earlier, this is a hard thing to prove in the Medical Examiner's office.

Unfortunately, this case also demonstrates that our system doesn't always work very well. By protocol, the child's body was taken to the Medical Examiner's office. The Medical Examiner's Office sometimes doesn't do everything correctly. In this particular case, I suspect the Medical Examiner did not look at the child's genitalia and rectum because they were listed in the Medical Examiner's report as normal. This made it difficult for us, as an institution, to deal with the District Attorney's Office because there was now inconsistent testimony as to what was observed. Fortunately we had some excellent photographs of the child, but we realized in retrospect, that we should have had our pathologist look at the body before it was transported to the morgue.

People may know facts about particular cases but are afraid to come forward. In this case a neighbor came to the emergency room with information, but refused to be identified because of fear of retaliation. This person told the registrar in the emergency room that this was a family well known in the community for intergenerational incest. Having this anonymous tip did not help in the prosecution of the case, because the informant refused to come forward.

The district attorney's office did the best it could. The other children were temporarily removed from the home, and this case went through both the criminal and family court system. It was hard to prove anything in criminal court because of the Medical Examiner's report.

The legal system is complex and sometimes cumbersome when it comes to dealing with cases of sexual abuse, incest, and child abuse. Although our laws do allow us to accomplish some things in the area of child abuse and neglect, unfortunately there are many children who still fall through the cracks.

Violence and Women: Case Presentation and Discussion

Stephanie V. Seremetis, M.D.

This case represents a typical episode of domestic violence.

A 34-year-old woman consulted me about fatigue and headaches. She is an accomplished, hard-working woman who was widowed at the age of 29. Her daughter is now 13 years old. The woman has been living with a man for the past 4 years. Her medical history was otherwise unrevealing, and the results of physical examination and laboratory assays were normal. She discussed at length her desire to give birth to another child. She expressed ambivalence about her current partner, both within the relationship and as a parent. Although I gave her many opportunities, she did not discuss specific issues. I encouraged her to undertake counselling regarding this important life decision and I offered her referrals.

Approximately 8 weeks after the initial visit the patient requested an emergency visit to discuss, as she put it, "problems at home." She explained that she had decided to try a new haircut and this had enraged her partner. He was verbally abusive and expressed what she described as paranoid jealousy. The next day, because she joined some friends after work, he again became enraged. This time the abuse was physical: he inflicted a number of blows to her head. She summoned the police, changed the locks on the doors, and notified friends and relatives of the events.

At the time of this visit, bruises were readily apparent on her face but there were no fractures. When asked if this sort of violence had ever occurred previously in the relationship, she indicated that, although it had not been part of the relationship during the past 2 years or so, it had been a significant part of the early relationship. During the past 2 years, her partner had continued to be verbally abusive. She described his verbal abuse as "merciless." He demanded that she curtail her social life and hobbies. She found that he saw any contact with friends or relatives as threatening.

She was reluctant to take time off from work to obtain a police order of protection, but she was strongly encouraged to do so. She indicated that she felt her most pressing need was for someone to talk to. I referred her to a counselor who was able to meet with her immediately.

Discussion

I would define violence as abuse or the unjust use of power. In this light, many, if not most women, have encountered violence in some form during the course of their lives. Violence affects the lives and health of women and has an impact on a physician's ability to take care of women and help them achieve the best health care for themselves and their families. Violence often occurs within the confines of a sexual relationship. We physicians must be careful not to limit ourselves to considering violence only in heterosexual couples; violence needs to be recognized as a part of some lesbian relationships as well. Of course, violence can occur in nonsexual relationships. We need to ask all women whether violence is a part of their relationships.

Violence can be a regular and frequent event, episodic, or something that occurs only once. Violence may be current, or may have occurred in the recent or remote past. When dealing with a woman, it is important to leave an opening in the

---

Associate Professor of Medicine, Director, Women's Health Program, Mount Sinai School of Medicine, New York, City University of New York.

course of the first few encounters for a description of prior events, including early childhood events. Certain such events can have health consequences for women and particularly can be at the root of depression, a condition common in women.

Detecting Violence

Violence in any of its forms must be detected. I believe that medicine has a bias against ascertaining whether violence has occurred. We physicians do not usually ask the question. In my practice, I try to ask a simple open-ended question: "Have you ever been in a relationship with someone who hurt you?" This allows the patient to understand that if the answer is yes, I am willing to listen. Also, it indicates to the patient that the issue of violence can safely be raised in a future encounter.

Violence against women is quite common and it is associated with significant morbidity and mortality. It must be considered a public health issue. Detection is probably best accomplished in the primary care setting, during routine screening. One must always ask the question in a manner that makes answering it possible.

Although the detection of violence is important, discovery is completely useless unless appropriate care or referral for care is available to the patient. If we raise the issue of violence and detect it in either a present or past event, we are not helping the patient unless we can then offer something further that will have an impact on the patient's life and health.

Legal Reporting

Unlike violence against children, physicians are not required to report cases of elder abuse or abuse of women. This is problematic and somewhat constrains us. Although it is true that most violence will be detected in the course of primary medical care, most primary care is not delivered in a setting that includes an established system to deal with the abuse. Most primary care is delivered in the office of an individual practitioner who is without an established or wide support network. Typically, resources for immediate referral and support are not available. Not having these proper connections can contribute to the primary care physician's reticence in asking questions about violence. This is a problem that requires systematic solutions. Women receiving care in a clinic often receive better care for the abuse they have sustained than do women who can afford to see a primary care physician in a private office.

One of the most disturbing aspects of domestic violence is that, although the victim may be helped to survive and attempts are made to protect her from future abuse, little effort is directed either at dealing with or changing the behavior of the violator or at dealing with and changing the system that perpetuates and supports the violators.

Violence in the Emergency Department

Sheldon Jacobson, M.D.

A 38-year-old African American man who has sickle cell thalassemia is a frequent visitor to the emergency department (ED) and the sickle cell clinic at The Mount Sinai Hospital for acute, as well as chronic pain syndrome related to sickle cell disease. His sickle cell vaso-occlusive crises that are painful bring him to our ED approximately once per month. On these visits he has typical pains in his spine and long bones. He usually is not febrile. His urine consistently tests positive for cocaine. In the past he has had aseptic necrosis of the femoral head and a cholecystectomy.

In the ED, he demands dilaudid or demerol intravenously for his pain. He will not take other opiates because of "allergy." The ED has a rule that after three doses of parenteral narcotics, the patient has to be admitted or discharged. On numerous occasions, this patient has remained in the unit for three doses of narcotics and then signed out against medical advice.

Because we do not give patients prescriptions for narcotics in the ED, he then contacts the personnel in our sickle cell program to fill the prescriptions for medication. He requests large numbers of pills. From informal discussions with personnel from the sickle cell program in three neighboring hospitals, we have learned that he also requests and receives prescriptions for narcotics from them. Recently, Medicaid has limited his prescription privileges to one Manhattan pharmacy to control this problem.

He has been a behavior problem for out unit. When he does not receive the services he is seeking in a time frame that meets his expectation, he becomes loud and verbally abusive. He moves about the room with escalating epithets and motor activity. This usually results in rapid attention to his requests and leads to the posting of security personnel at his bedside.

During one of his recent unscheduled visits, the patient was accompanied by his 7-year-old son. He indicated that his last prescription, given to him by another hospital, was refused by his pharmacy. He apparently had forged an increased amount of medication to be dispensed on his prescription. The pharmacist recognized this forgery and turned the prescription over to the police. The patient asked that a new prescription be filled out and noted that he was having acute pain. He was referred to the ED for his painful syndrome. After leaving the clinic, the patient had what appeared to be a seizure in the hallway and was surrounded by numerous solicitous and supportive hospital personnel. When members of the sickle cell team arrived on the scene, the patient stood up immediately and attempted to rush off. By this time, there was a serious question about the well-being of the patient's son, so a hospital security officer at the scene was asked to detain the patient for further evaluation. At this point, the patient reached into his pocket, withdrew a pencil, and lunged at the security officer with it. The patient was subdued and taken to the psychiatric ED. The patient's son was cared for by a hospital social worker until the Bureau of Child Welfare came for him. Eventually the patient was discharged to his home for the night.

Discussion

This patient is obviously a troubled individual. He has several serious organic diseases that present a number of management problems for our staff. In each of these problematic areas we are confronted with ethical dilemmas:

Professor of Emergency Medicine, Chairman, Department of Emergency Medicine, Mount Sinai School of Medicine, New York, City University of New York.

- What is the extent of our responsibility to care for this patient who represents a danger to our patients and staff?
- To what extent can we force the patient to live up to a prenegotiated contract to use only one hospital and to attend the clinic regularly in order to continue being cared for at the Mount Sinai Hospital?
- During his visits a number of diagnostic tests are performed and one-on-one security is provided during each of his stays in the ED. Can the principle of distributive justice be invoked to limit our resource allocation to this patient?
- Can patient confidentiality be overridden to pool information from regional hospitals about this type of problem patient?

Violence and the Physician's Duty:

A Panel Discussion

DANIEL A. MOROS, CHAIRPERSON

Moros: Witnessing sometimes involves a responsibility to act. In medicine, arguments have been made in favor of reporting laws that would require physicians who witness violence to act, report, and intervene. To what extent does our commitment to individual freedom and patient autonomy compete with and limit what we, as physicians, should ask for in terms of the power to intervene?

Givelbar: What you are asking for is something akin to the Tarasoff decision, a legal rule, imposed from above, which you don't have to seek actively. Such a rule would dictate that, in the case of suspected elder abuse or domestic violence, a physician is libel unless a report is filed. Part of the power of the Tarasoff case was that nobody in the treating community had any specific responsibility for reporting. The rule came into law obviating any ethical dilemma about what physicians ought to do. I think it is a fair reading of New York law, right now, that if you believe someone is in danger of violence by a third party, you are not violating any law by taking action to reveal the possible danger to the appropriate authority. Still, a feeling remains that ethically physicians should not do this. It appears that people would like a law that compels reporting.

Gert: I'm not sure what the arrangements ought to be, but it is clear that there is some latitude in what they can be. It would seem that if we could agree on what might be required, everybody should know what they are including patients and hospital personnel. Any policy with regard to confidentiality and when it may be breached has to be a public policy. If this is not the case, the principle of informed consent is violated. Before people come in for help, they have to know there is the possibility that what is found will be reported. This seems noncontroversial. If there is no system in place to do something once an incident is reported, then it's not clear what the point of the reporting would be. So another noncontroversial point is that the system needs to be improved so victims of violence will benefit by its being reported. If there is no system, the people are forced to go back into their original situation and might well be worse off than they were before.

Bar On: One of the issues that comes up in a discussion about reporting where competent adults are involved has to do with who should decide what is in the best interest of the patient. The problems with trying to solve an ethical dilemma like this one are social and economic. If we are going to take people out of their homes, out of their present situations, where will we put them? We think we know better, and so we take them out. But we are not solving their problems. We don't have solutions to their situations. We need to listen to people carefully and worry, not just about their survival, but also about their dignity. We need to be careful that all of the arrangements we plan will preserve their dignity.

Eiser: I think the cases illustrate that we have to listen not only to the individual patient, but to pay attention to the larger social forces that are creating these situations. To try to solve these problems on a case-by-case basis will not resolve the larger problem. If a social or cultural structure inhibiting violence is missing, then we cannot begin to solve problems case by case.

Jacobson: I want to raise the issue of the rights of health care workers. I think we should give a patient as much as we can. But what are we to do when one patient is usurping a majority of the resources and at the same time threatening the health of other patients and staff? In the case that

I presented, the person has been put on Mount Sinai's special patient list and when he comes to Mount Sinai a security guard accompanies him throughout his stay in the hospital. Perhaps that keeps us safe for the time being. But he could appear without being recognized, or use an alias, and we are again in danger. He also receives medication which he sells in the community and so we are in complicity with his drug peddling. There must be a point beyond which we are not required to go. I'm not sure where that is. Mount Sinai has decided that this is a potentially explosive situation. They try to control it as best they can through security measures.

Givelbar: Perhaps one way of approaching all these problems is to assume that the law permits you to do anything. I think there is a tendency to use fear of litigation as a reason for not acting and to glorify the danger of litigation as a reason for not acting. There should be clarity about whether one is worried because the law says you can't do something, or because there is a sense that it is unethical to do it, or because one's responsibility to the patient prevents one from doing something. Sometimes, imagining that the law plays no role might allow one to get closer to what one actually ought to do.

Gert: Given some conception of autonomy and some limits, if the patient is told and knows what will be done when he comes in, then one can do it. For example, if a patient is informed that when he shows up in the emergency room for drug therapy, the records of other hospitals will be checked to see if he is getting drug therapy from multiple sources, then he is aware that this is the policy. It's then fine to check with other hospitals.

Bar On: Judging from this case, the patient is an addict and so to say that informing him takes care of everything isn't quite right.

Gert: What I meant was, if we address the question about whether or not it is wrong to call other hospitals to find out if the patient is getting a lot of medication from a lot of hospitals, we can operate in a way that is not wrong. If we inform him that we are going to keep tabs on him to make sure he doesn't exploit the system, then what we are doing is not wrong. As long as he is told beforehand that this is what will happen, we can do it.

Moros: Dr. Lifton, it strikes me that the experience of witnessing violence elicits a tremendous desire on the part of physicians and health care workers to respond, but it is not immediately clear what a coherent response on our part, or the hospital's part, or society's part ought to be. The witnessing certainly makes us uncomfortable and desirous of responding.

Lifton: What I hear from the presentations and the panel of case presenters is that medicine, doctors, and all medical workers are being caught right in the heart of society's ills. We speak of an epidemic of violence and it is a general impulse of medicine to apply itself to this violence, but there is no existing medical institution able to do that. Witnessing means opening oneself to the experience. It is, to not do what many do, not to push it away or under cover or to deny the experiences of abuse and violence, but rather to record them. It may be that at this stage medicine can do less in terms of treating violence than it can in bearing witness to violence. That's what doctors in the antinuclear movement did. Our aim was to point out that we cannot patch people up in the event of a nuclear war. There was a so-called bombing run where we would explain what would happen if a one-megaton bomb were to land in the middle of the city. It was a way of imagining right in one's home area what the harmful technologic instrument of genocide would do. What is the parallel here? Children are being killed by parents or other adults, elders are being killed or abused, often by their own children, stealing their money or physically abusing them. It may be that doctors with direct experience of these kinds of violence have to start organizing and presenting this material in both medical and nonmedical forums, in public meetings or public statements. To do that, one has to probe further into the causes of the violence that one encounters, into families and subcultures and communities, and to make it known. Physicians need to cooperate with community groups as much as possible not only to treat the victims medically but to transform their private witness into social knowledge.

Paris: From my own experience over the past 10 years, there are two areas in which we can make some headway. Violence has been part of society forever and no doubt will continue to be. We can attack the problem by initiating two interventions. First, we don't teach students in medical school how to talk to patients, how to interview patients, and simply to ask in the interview whether or not anyone harms them. It should be an essential part of an assessment along with things like asking how they get their food. If people are not asked this sort of basic question, we will not identify those patients who need special help. Second, someone needs to deal with the perpetrators. The physicians don't and can't deal with the perpetrators. Sometimes we can, for example, give our patient references for rehabilitation counseling for their addicted child who is stealing, but that is the extent of the physician's contact with the perpetrator. What really comes across is that this is a domestic or family matter. In the case Dr. Jacobson presented, I'm worried about the patient's 7-year-old child who is often in his company. We need to look at an individual family for assessment. We need to train those who will come into contact with violence to deal with it, as a practical measure, regardless of reporting.

Baumrin*: In reference to Dr. Eiser's talk, an interesting fact is that when the circumcision and or infibulation is carried out, the girl's mother helps to hold her down while it is being done.

Eiser: The process is generally done in a very primitive fashion, often performed by women, and the mother is involved. The feeling is that without going through this process, the girl is not a fulfledged female in that society. And so, the mother willingly goes along with it.

Gert: I think one of the things this shows is the power of culture and ideas. If you look at the atrocities worldwide, you realize the power of culture to overcome natural revulsion is truly amazing. Thomas Hobbes, who was referred to in Professor Bar On's paper, compared man to animals, saying that we are capable of much better things than animals but also capable of doing much worse things than any animal would do.

Bar On: I'd like to point out that in all cultures women are entrusted with raising the next generation, especially the next generation of women. In the case of female circumcision, it's simply the cultural process of the production of women in that society. And so, in that context, it makes perfectly good sense that women circumcise their daughters. Women in our society are entrusted with making the women in this society. Women are told what to do by women, for example, when to shave their legs. It's just part of the production of the proper culturally gendered person. Women are entrusted with that job.

Lifton: There are certainly some societies which are less violent than others. One wants to look at what in their institutions mediates violence because there will still be some violence and violent impulses. There must be some sort of mode of realization that replaces violence and that is worth studying.

Rabin*: I don't really see how there can be witnessing, reporting, and action, because that is in direct conflict with respect for autonomy. It seems that there is tremendous conflict and I worry that there will be more rather than less violence in the future.

Givelbar: Believe it or not, in some ways the situation is much better today than it was in the 1960s. The witnessing and the naming are important, and perhaps right now one can't do much more than witness and name. In the 1960s I was given a manual that said whereas middle class couples may spar verbally, many of the people we would encounter would be, in essence, preverbal and, therefore, they would be whacking each other. This was accepted as true. The rule in the office was to not try to prosecute cases of domestic violence because the women wouldn't show up. It was believed they didn't show up because the physical abuse was just their way. An incest case, in those days, was something you referred to counseling. It was not viewed as it is today. This doesn't mean we presently have any idea of what to do about the different forms of violence. The criminal justice system is making progress and people are becoming conscious of it. I don't think it's getting worse, I think it's just that it is much more visible because it is identified. I agree with Dr. Lifton, witnessing is important. We've all been through an era where violence wasn't witnessed, it was not acknowledged, it was thought of as background noise.

Lifton: I think from our own personal standpoint we are susceptible to frustration and all sorts of difficulties in trying to carry through what seem impossible, all the more so because many of us have a conscience. One has to recognize that one has limitations. One has to find the niche with other like-minded people in trying to do something, perhaps with a fairly humble goal or set of goals. One has to keep witnessing and deepening one's grasp of what one is doing and making that information known. While they may not be the satisfaction of immediate results, one is potentially contributing to long-range improvement. Individually and together, there is still room to improve.

Bar On: I'd like to note that important social movements occurred between the 1960s and now. Many concerns were expressed in those movements. The reason we now cannot find solutions for certain individual cases is because they require that kind of work—the work of social activists. An enormous educational effort is necessary in this area. Education will eventually change things.

---

*From the audience: Bernard Baumrin, Ph.D., Professor of Philosophy, Lehman College and The Graduate School, City University of New York.

*From the audience: Estelle Rabin, Ph.D., Assistant Professor of English, Queens College, City University of New York.

Eiser: Even in bioethics there has been movement in the area of violence. It is a slow process and requires work. At the Hastings Center there is an ongoing project on the goals of medicine. The fact that violence, in general, and domestic violence, in particular, makes a great impact on medicine, keeps coming up. The question is, where does the physician's responsibility lie with such things? There is agreement that the physician's ability to intervene and curb violence is limited in this society.

Robichaud: Because Dr. Jacobson works directly with patients like the one in the case he presented, I would like for him to say more about what he thinks possible solutions to these sorts of cases might be.

Jacobson: I think sometimes it may depend on how the patient's health care is funded. It seems to me that different issues are posed when the patient's health care is supported by public or private funds. In the case I presented, the patient is receiving medical assistance so we are funding some of his sociopathy. What I would like to do is to convene a panel of interested and knowledgable people, including a patient advocate, sickle cell anemia experts, legal experts, and develop what we think is a reasonably safe and effective way to manage this patient. We would discuss it with him and set up a binding contract with him. If he tries to go to other institutions, there should be a mechanism in place to stop his abuse of the system. For example, his medical card should only be accepted by other institutions in the instance of a life-threatening condition. If it is a routine sickle cell presentation, he should be required to go to one place, have one system, to get no prescriptions elsewhere and so on. He has three diseases—sickle cell anemia, drug addiction, and probably a sociopathic personality. It would be very difficult for him to live with that sort of contract. Thus, in his case, it might be an exercise in futility. He's the type that would present with chest pain, a potentially life-threatening condition, then within a few hours would be requesting pain medication. In other cities, I tried this method of engaging the patient in a contract. It only worked when patients remained at my institution.

Rhodes: It sounds as though we moved beyond a very constricted view of what the duty of health care providers is to a very expanded sense that can even encompass asking patients where they get their food. It seems to me that this is a real extension of the concern for the patient's good. We can make it easy for women to answer questions about violence. We can create women's health programs which invite women into a different kind of environment and have these questions addressed. My question is, is it appropriate for medicine to move into this broader social domain.

Paris: Yes. In the case of the young woman presenting with a headache and fatigue, those conditions were not what was really bothering her. However, unless the question is asked, there is no way to know and one can't then treat the entire patient. For example, getting enough food for elderly patients is a real problem in New York City. If the patient is not asked and the problem of food procurement is not solved, then the patient will be malnourished, and so, not healthy.

Seremetis: There's really no magic to this. These are simple interventions. It's a shame they are not taught in medical school or during medical training. It is also a shame that both male and female primary care practitioners will deny that domestic violence is a common problem in their practice. They think that because their patients are middle class, domestic violence is not going to be a problem. If you don't ask the question, often you won't find out. I think it's important for our profession to witness and to talk about the experience of how simple things like providing a comfortable environment allow patients to reveal these sorts of problems to the physician.

Grimm: This does broaden the boundaries but some of these social problems, which involve abuse and lack of nurture and lack of love, are complicated problems with no easy solution. There are a lot of bright and gifted people who work in these areas, yet there is still a lot of spinning of wheels. One approach which has been shown to be effective is an interdisciplinary approach, one which we use here in East Harlem. We use a psychologist, a social worker, a teacher, a clergyman, each with a different vantage point. Some of the best program planning and efficient use of professional time has involved this network. In the area of child abuse, because there are laws governing advocacy, the only way to really work is to use an interdisciplinary approach. We try to coordinate the multiple services involved, because that is also best for the family.

Epilogue

Rosamond Rhodes, Ph.D.
Associate Professor of Medical Education
Director, Bioethics Education
Mount Sinai School of Medicine

Dr. Barbara Paris's case of elder abuse and neglect

Elder abuse and neglect may be difficult to detect. Frequently the patient is embarrassed by the situation and collaborates in concealing the problem. Ageism, the isolation of the patient, and physician reluctance to confront and intercede all undermine medicine in facing these important problems.

Dr. Katherine Teet Grimm's case of child abuse

Children die as victims of abuse. Even those who survive suffer significant physical and psychological harms as a result of the violence inflicted upon them. The social service network, which has the responsibility for dealing with these cases, allows many child victims to fall through the cracks.

Dr. Stephanie V. Seremetis' case of violence against women

Violence against women is common in our society. It is frequently associated with significant morbidity and mortality. As part of routine care, doctors should be asking the question, "Have you ever been in a relationship with someone who hurt you?" The primary care environment should make it possible for patients to answer that question honestly. Furthermore, the primary care physician must have access to a support network for addressing the problems that are revealed.

Dr. Sheldon Jacobson's case of violence in the emergency department

Violence by patients cannot be tolerated in the health care setting. In dealing with violent patients the physician confronts ethical conflicts about the predominance of the duty to act for the patient's good and the duty to preserve confidentiality. Because containing violent patients and limiting the harm that they can inflict places increased demands on the health care system, providing them with medical care raises the issue of how our limited health care resources should be justly distributed.

Mount Sinai Medical School Commencement Address
May 19, 1995
New York City, New York

Antonia C. Novello, M.D., M.P.H.
UNICEF Special Representative for Health and Nutrition

It is a great honor for me to be a candidate for the Doctor of Humane Letters, honoris causa from Mount Sinai School of Medicine. I am humbled to think that now my name will join the roster of so many extraordinary men and women who have been honored by this outstanding medical school. I will do everything in my power to live up to the spirit and commitment that goes hand in hand with this honor.

As past recipient of your Community Service Award as well, suffice it to say, I am extremely impressed with all of Mount Sinai, the students, the faculty, and all that you stand for—and stand behind. For the past 32 years, guided by the most honorable, fundamental principles, this School of Medicine has truly served the medical needs of this city as a "school without walls." If, as some say, "your campus is the community," then you are truly some of its finest leaders.

Regarding your class, I am impressed that there are 138 graduates, 74 men and 64 women; a large number of you are graduating with honors, and 23 of you have been elected to Alpha Omega Alpha. It comes as no surprise then, that nearly all of you matched your first and second choice of residency programs and, as a consequence, a good number of you will be going to very prestigious institutions. I am equally impressed that some of you have found time to get married and start families in the midst of learning how to care for others. Equally as impressive, many sons and daughters of physicians have decided to follow in their parents' footsteps. Congratulations, it's extremelyrewarding to see this!

As a minority, I am even more impressed that a very large number of you will graduate and will serve as an inspirational example to the people of this nation. For that, I thank Mount Sinai and congratulate all of you. The American dream is alive and well.

That is why, as a woman and as an Hispanic, it is an honor for me to join you today in celebrating this outstanding institution and to share your pride in these fine young men and women who will carry your legacy into the future—your graduating class of 1995. Your class is small enough to be personal, yet large enough to be powerful. Graduates, I stand before you today to do three things: to *congratulate* you, to *counsel* you, and to *challenge* you.

Congratulations. Just think of it—in one month you can finally start paying off your student loans. I do want to congratulate you on your entry into the profession of medicine. Yes, you are now back at the entry level. A cliche? Sure it is—but, like most cliches, it contains a kernel of wisdom—especially for physicians.

Think about it. Most of the science and technology you have learned will soon be obsolete. Even worse, in 10 years you probably won't remember more than 10% of all the facts you have had crammed into your heads.

It troubles me that so much of medical education must be spent in the drudgery of memorizing *facts,* and so little time spent in the conceptual teachings of public health, prevention, and common sense. Facts are necessary, of course. They are the building blocks of our knowledge.

Without facts, scientific judgment is *impossible,* honesty is *irrelevant,* and compassion is *fraudulent*—however, facts change. That is why, for me, the true purpose of a medical education is not merely the accumulation of *facts,* but the creation of a *habit of mind, a way of being.* By this I mean: The complete health professional must be humble in what facts he or she does not know.

Socrates said that if he was wiser than all the rest in Ancient Greece, it was only because *he knew what he didn't know.* That is why I prefer to evaluate health professionals not so much on the *answers* they give, but on the *questions* they ask. So, while I congratulate you most warmly today, I urge you to embark upon a lifelong quest of learning, of healing, of caring . . . but, above all, of *questioning.*

You stand upon the shoulders of your faculty, just as they stood upon the shoulders of their predecessors. And, like them, you too will one day pass the torch to a new generation. Graduates, you must keep the torch of medicine burning brightly! So cherish and advance your profession. In the words of Hippocrates: "Whenever the art of medicine is loved, there is also a love of humanity."

And now, my counsel. Sir William Osler, in his farewell address to American medical students, said, "To each one of you the practice of medicine will be very much as you make it—to one a worry, a care, a perpetual annoyance; to another a daily joy and a life of as much happiness and usefulness as can well fall to the lot of man." And so I counsel you to set your goals. After all, if you don't know where you're going, you're already there!

What kind of physician will you be? I want you to be known not only for your triumphs, but also for your dreams. So plan thoughtfully and realistically, but dare to reach for the stars. Although I pray that you will be known by your successes, remember that there is no shame in noble failure. We must remember that our work as professionals defines us as human beings. As it has been said, "much of the good and the bad you will ever do in this world will come through your work."

How will you view your work? The author Robert Bella describes three types of work. First, there's the job, where the goal is simply earning a living and supporting your family. Then there's the career, where you trace your progress through various appointments and achievements. Finally, there the calling, the ideal blending of activity and character that makes work inseparable from life.

My friends, I hope you are not just looking for a job. I hope you are not just planning a career. I hope each and every one of you has a calling—a physicianly calling. I know that you will do well. Your faculty has seen to that. But I pray you will do good.

One word of caution, though—from one who sat where you are—25 years ago, *the world owes you nothing!* To expect the world to treat you fairly because you're a good person and a graduate of Mount Sinai is like expecting a raging bull not to charge you because you're a vegetarian! So work hard, and beware those enemies of tranquility: avarice, ambition, envy, anger, and pride; and avoid Kuschner's pillars of despair: complacency, mediocrity, and above all, indifference.

Will you be happy in your work? I hope so. But how will you do it? Riches, fame, and power alone will not make you happy. In fact, the happiest physicians and health care professionals I know are those who go about their everyday lives—doing good things for other people—without asking "what's in it for me?"

We must remember, as it has been said, that service is the rent we pay for living . . . and that service will set us apart . . . service to *ourselves* and our *families,* service to *country* and service to *God.* The importance of this service attitude is painfully obvious in today's world. My friends, there is a spiritual vacuum in America's society today. There is too much of "what can I get?" rather than "what can I give?" We must remember that the test of making a living is how much you get; the test of making a life is how much you give.

This is why, while I fervently wish you much happiness in your work, I caution you not to seek happiness for happiness' sake. If you do good, happiness, believe me, will seek you. It will creep silently into your lives from many sources: the heartfelt "thank you" from a grateful patient, the quiet discovery in a laboratory, the well-crafted lecture, the scientific paper, the book, or the public policy that may affect the lives of millions. Medicine, you see, offers many paths to happiness. So I counsel you not to be good doctors or good health care professionals only. First and foremost, be good men and good women.

Do not forget to put balance in your lives: family, friends, leisure time, hobbies, social causes, volunteerism, and religion. Only by maintaining your own health—physical, mental, and spiritual—will you be able to look after the lives of those entrusted in your care.

Now, I would like to challenge you. Colleagues, I don't need to tell you that the future of public health is paved with complex issues, but I can certainly tell you that the future is already here, and much remains to be done. I am reminded of some very simple, yet profound observations of Mark Twain, who said, "Good judgment comes from experience. And where does experience come from? Experience comes from bad judgment." I can tell you that if we are going to meet the Nation's complex and difficult public health challenges we're going to need healthy doses of both experience and judgment.

The truth is that we are witnessing a staggering epidemiologic, demographic, and political revolution and, like it or not, we will all be swept into its currents. I believe that, in order to succeed, we must rediscover the true practice of health care, which is much greater than the treatment of illness alone, or we will be practicing health care and medicine in a vacuum. I also truly believe that if we are willing to right the wrongs of oppression, injustice, famine, and poverty around the world, we must be willing to extend our compassion around the corner right here at home.

In this gallant quest let us not forget those without a voice in our society: our children, our poor and homeless, our immigrants, and our aged. And let us not forget those unfairly stigmatized by AIDS, physical disabilities, and mental illness. Colleagues, although it may be impossible to determine precisely what the changes will be, allow me to share with you five challenges that I see along the way, challenges that might have a direct effect on the way we deliver care and, in addition, might have an impact on the way we respond as a medical community.

Challenge number 1. We must respond to the changing demographics in this diverse nation of 250 million people—with more than half of the population women, with close to 15% of our population African American, more than 9% Hispanic, 3% Asian or Pacific Islander, and 1% Native American; with almost one-in-five of our citizens being younger than the age of 12, and more than one-in-ten older than age of 65—we simply cannot afford to cling blindly to values and perceptions that are out of touch, or to adhere rigidly to outdated practices accepted by the mainstream. We need to recognize that the role of the health care professional will have to change to accommodate the shifts in demographics and the special needs of emerging communities.

I believe that today, as never before, our diversity as a nation is intertwined with our strength and our hope for the future. Based on this strength, and bound by our common concern for the health and well being of all people in this nation, we can join forces as never before to close the gap between the "haves" and the "have-nots." If America is to be great for any of us, then America must be great for all of us. Franklin D. Roosevelt said, "The test of our progress is not whether we add to the abundance of those who have much. It is whether we provide enough to those who have little."

Challenge number 2. The second challenge threatens to shake American medicine to its core: the challenge of access for all our citizens to adequate health care. Care that is comprehensive, family based, community centered, culturally sensitive, and under one roof.

We need a health care system that is user-friendly, portable, efficient, and secure, and meets the five A's—*accessibility, affordability, availability, accountability,* and, to be sure, *affability* as well; a health care system in which there is an appropriate balance between meeting the health care needs in a cost-effective manner, without jeopardizing the number of physicians, the access to the health care provider or where teaching and or research are curtailed to an unnecessary degree. Obviously, an ideal plan for today's needs.

Many of us in this audience can remember the mid-1970's, when the public held physicians in high esteem, right next to members of the clergy. Confidence in the ability to make a decent living was always assumed; all we had to do was to "hang up a shingle" and patients would come. Academic medicine seemed to offer the best of both worlds: plenty of time for research, some teaching, and limited practice responsibilities.

Today, it all seems to be jeopardized somehow: the independence, the time, the easily acquired grants, the life-style, even the security. Control has been wrestled from medicine and is in the hands of bureaucrats, CEOs, and even politicians. And as Dr. Steven Wartman has said:

> The image of medicine that is being portrayed is the one in which doctors are more concerned about protecting their own incomes and personal pursuits, than they are about society as a whole. The price that we have paid, has been a continuing loss of independence, credibility and a fear for the future. While it might be tempting to look with disdain at what is happening in health care reform around us, it might be a great mistake to do so.

After all, we might not be able to control all that is happening in the political world of today.

What options will be open to us physicians in this kind of environment, you might ask? The usual political options, such as advocacy and public education, are dependent upon physicians agreeing as a group upon a common approach. Unfortunately, you will come to realize that such accord among us as a group is highly unlikely. We physicians do not have a tradition nor the time for becoming trained politicians. We take care of people—not polls!

I believe that physicians in their practices, and medical students in their career choices, must not depend on politics alone. Instead, if we are to succeed, we must start to respond more directly to the needs of society to regain some semblance of control over what is happening to the health care system and, as a consequence, to our future. After all, as Dr. Wartman has said, "Whether administrator or surgeon, internist or pediatrician, radiologist or anesthesiologist, you will not lead our country into the 21st century regarding health care, if you operate with tunnel vision, if your administrative ability is bounded by red ink, if your surgery reveals simply a superb artist, if your eyes cannot see beyond the immediate cause of malnutrition or a gunshot wound." I submit that your God-given vocation goes beyond your ability to cure.

May I suggest then, that irrespective of what the future of health care might bring, that you, the health care leaders of the future, become men and women who are concerned with discovering where your patients come from, and where they've been. Not only their medical history on a reception-room questionnaire: but their broader experience of human and inhuman living—in reality, their total persona, their integral humanity. "Remember, your patients will not care how much you know, until they know how much you care."

Challenge number 3. My third challenge is one that is dear to my heart as a pediatrician. I'm talking, of course, about America's children. It has been said that if you wish to take the measure of any nation, you need look no further than how it treats its children.

Today, we know children are more likely to fall below the poverty line than Americans in any other age group, including the elderly. At any point in time, roughly one in five children younger than 18 years and one in four children younger than 6 years lives in a family with income less than the federal poverty threshold, for a family of four. No other industrialized nation has anywhere near this level of child poverty.

As we search for answers, let's not forget to ask the right questions. And let's not pass the buck. Contemporary American epidemics will not be cured by simply throwing money at them or by packaging new federal programs. Society's problems are medicine's problems. Children generally do as we do, not as we say. And values are much easier caught than taught. So let's start teaching by example.

Remember, children do not come in a vacuum, children come from families. Society, however, has placed families in rigid categories and has created, in turn, negative terms to describe family systems that nonetheless do possess strengths even in the presence of turmoil. If we are to succeed, we must purge family structures of gender bias and experiment with new relationships within families, and with new forms of families attuned to the changing social context of the world today. And we must empower all families—whatever shape they may have—to nurture their children.

As we plan for the future of our children, let's plan for all the children, not just those who receive political attention. In the words of the poet Ina Hughs:

> Don't forget the children who never get dessert, who have no safe blanket to drag behind them, who watch their parents watch them die, who can't find any bread to steal, who don't have any rooms to clean up, whose pictures aren't on anybody's dresser, whose monsters are real—and whose nightmares come in the daytime. . .

Don't forget about the children.

As we approach the year 2000, I also ask of you to not forget about the women. How can we afford not to take the lives and future of women into our hands, and secure a better place for them in the millennium ahead. After all, we are 127 million strong, 29,000 of whom are alive past the age of 100, we represent 46% of the nation's labor force and, even when single, we take care of 25% of America's households.

From time immemorial, women have taken care of their families. But when they are sick, who will care for them? The time has come for women to take care, not only of their families, but also of themselves. That is not selfish; to the contrary, that is absolutely life-affirming. So let's put all the guilt and blame and self-doubts aside, and embrace the truth: when the health of a woman collapses, the health of the family collapses as well. *So don't forget to care for the women.*

Challenge number 4. The fourth critical challenge is AIDS. Today AIDS is the first leading cause of death for all men between the ages of 25 and 44, the fourth leading cause of death for all women between the ages of 25 and 44, and the eighth leading cause of death for all Americans. And despite this, there are too many who still believe that only homosexuals are at risk. In this second decade of AIDS, homosexuals are not the only group at risk; children, adolescents, and women continue to be at increasing risk as well.

We must therefore find new, more feasible ways to deal with the enormous public health problem posed by the HIV/AIDS epidemic. Indeed, our primary concern must be for the provision of education for all people in this country regarding HIV/AIDS.

There are entire families at risk, and we are missing the window of opportunity to make an impact on their lives. In this second decade of AIDS, I am afraid that some of us are still assuming, denying, and projecting. And more importantly, we are still judging—rather than caring.

With increasing heterosexual transmission of the AIDS virus, the HIV epidemic will undoubtedly pose the greatest health challenge of our generation. The spectrum of health care needs for HIV-infected patients will challenge all our creativity and imagination; and for this, we must be prepared. Above all, we must be prepared to fight AIDS as much with our research as our resolve. In the future we must care more and judge less.

Challenge number 5. Finally, challenge number five is another that I have addressed as your former Surgeon General: the challenge of violence in our society—a threat that is unraveling the fabric of society, casting a shadow in every home and community, and robbing our peace of mind. Physicians are used to thinking that the issue of violence is the responsibility of other groups—the police, the courts, politicians, the media, and the church. I would suggest, however, that this thinking needs revision. Family violence is a disease, but for way too long it has been treated like a long-held family secret. Until we start talking about it, until we start treating it like the disease it really is, it will continue to devastate families and destroy lives.

Colleagues, it is no small problem that: one in four American women is affected by domestic violence; every 12 seconds, 24 hours a day, 365 days a year, a woman is punched, kicked, slapped—or worse; every decade, violence *kills* as many women in this country as the total number of Americans who died in the Vietnam War. Imagine what it would be like to build a new “Wall” every 10 years, a wall inscribed with the names of American women who have died, most often at the hands of those who should love them most. But, unlike the Vietnam memorial, this imaginary memorial keeps growing—one slap, one punch, one bullet at the time.

As horrible as the numbers are, this hidden war doesn’t just affect women. Domestic violence targets the most vulnerable among us, men, women and children, including those least able to help themselves and those at the beginnings and ends of their lives. While there is not enough statistical data to be absolutely accurate, we do know this about the impact of violence on medical services:

- Every year, domestic violence results in 30,000 emergency room visits, 40,000 visits to a physician, and almost 10,000 days of hospitalization.
- Each year, the direct medical costs of all violent injuries in the United States leads to an estimated $10 billion in health-related costs, $23 billion in lost productivity, and almost $145 billion in reduced quality of life.

If the economic costs of violence are high, the social costs are incalculable. The truth is that we health care providers at times have failed to properly identify as many as 95% of the victims of domestic violence. Although we providers spend our lives on the front lines, where we see the victims every day, sometimes we do not see them, sometimes we do not listen. Sometimes we do not respond to them. Remember, the security of a physician’s office may be the only sanctuary a patient can find in his or her personal hell of violence and neglect.

The reality is that you and I may be all that stand between the patient and continued abuse. So learn to ask, learn to document, learn to refer, learn patience. Most important, learn to understand.

Physicians, as the guardians of lives, must speak up. We must get informed, get involved, and get in charge. I maintain that violence will only stop when all of us stop accepting it as a way of life and begin to speak and act as part of the solution.

So, there you have it. A *top-five list of challenges* from your former Surgeon General. A “hit list” if you will, but of the healthiest variety. The going will be tough, but as I look out at the Class of 1995, I am encouraged. Sitting before me I see our new generation of clinicians, teachers, researchers, communicators, administrators, and perhaps, even the next Surgeon General. I believe you are ready to meet these and all the other challenges that lay in wait for you.

To effectively address the five challenges I have outlined, you must learn to *work effectively* with other members of the health care team—for in the medical world you inherit, you will no longer be the sole proprietors of the state of America's health. And, you must also learn to *communicate effectively* as you will become educators of your communities. You will be a guide, a voice, an advisor, as well as a person of considerable power, in helping patients make correct decisions. So, speak up and *listen* well. Believe me, you will listen your way into your patient's lives much more meaningfully than you will ever talk your way in.

Colleagues, before I close, I would like to pay special tribute to your parents. In the recent rush of intern matching, final examinations, moving out of apartments, and yes, the endless round of celebration with your friends, please remember that this is your parents' day every bit as much as your own. While you can never really know what it is to love a child as a parent does until you yourselves are parents, I hope you will find a special moment sometime later today to express your gratitude, and your love, to them. There can be no eluding your responsibility to give something back to your parents, your community, your faculty, and all of those who helped make it possible for you to be here today.

Today, it has been my distinct honor to congratulate, counsel, and challenge. The rest is up to you. May you seize this day and all others that are to follow to bring honor to your alma mater, joy to your family and friends, comfort to your patients, and true happiness to yourselves. Above all, I pray that you do not lose your sense of who you really are, where your roots are, where you came from, and who helped you in getting to this point. Most importantly, do not forget the impact of this great institution in molding your life and professional future.

Mount Sinai's Class of 1995, may you think clearly . . . act decisively . . . and care tenderly. For without knowledge there is no future, and without caring, there is no hope. Thank you, and God bless.

The Heart and Thyroid Disease

Paul W. Ladenson, M.D.

Cardiovascular Complications Of Hypothyroidism

A Young Man with Heart Failure

I would like to begin by presenting to you the case a 19-year-old man who came to the Johns Hopkins Hospital in 1990 with biventricular heart failure, a university student who had been referred in anticipation of cardiac transplantation (1). He had a 2-month history of progressive fatigue and dyspnea on exertion, and a 4-day history of pleuritic chest pain. Soon after his arrival at our institution, it was determined that he had an ejection fraction of only 16%. He was a very sick young man.

When we looked at some of his old school photographs to try to reconstruct what had happened to him, they gave us unique insight into the natural history of his disease (**Fig. 1**). Ten years before admission, he looked rather healthy and was, according to his mother, active—participating in soccer and other sports—and doing well in school. Few of us would note anything abnormal in this young man, although one of my astute thyroid colleagues commented on the fact that he seemed to have had mild infraorbital edema even then. In the 6 months between March and October, 1982 (8 years before clinical presentation), you can see that something had occurred; there was rounding of his face, weight gain, and obvious periorbital edema. By the next year, it is clear in retrospect than this young man was hypothyroid. Despite the fact that he came from a family with good access to medical care, his condition was not suspected. During these years, he was once evaluated for iron-deficiency anemia but, perhaps because he was a boy, his pediatrician did not consider thyroid disease. At the time of his admission to our hospital, he was clearly myxedematous with sallow dry skin, dry hair, a docile, somewhat apathetic disposition, a little loss of the lateral eyebrow, and marked periorbital edema. His initial serum thyroid function tests confirmed the presence of severe hypothyroidism with an undetectably low serum-free thyroxine, and a serum thyroid-stimulating hormone (TSH) concentration 794 mU/L.

The Myxedema Heart

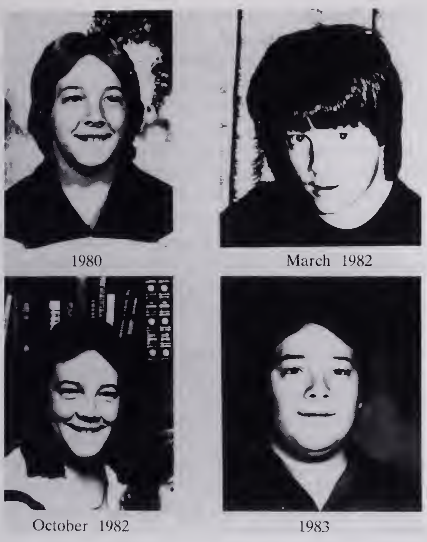
This young man with profound hypothyroidism and severe biventricular heart failure was certainly not the first such case in the annals of medicine. The first report of the myxedema heart was made in the German literature in 1918, when Herman Zondek reported his observations of hypothyroid patients with heart failure that was treated with desiccated thyroid (2). Zondek described these patients' initially widened pericardios cardiac silhouette, slow and indolent cardiac contractions, and low electrocardiographic voltage, all of which were reversed by treatment. He observed the kind of improvement that you see in this modern pair of chest x-ray images of a patient with hypothyroidism, before and after treatment (**Fig. 2**). Incidentally, Zondek's trial was controlled, in that he also administered thyroid hormone to patients with rheumatic heart disease causing heart failure; and they, of course, did not get better.

Pericardial Effusions in Hypothyroidism

Much of the widened pericardiocardiac silhouette in hypothyroidism is caused by a common feature of hypothyroidism affecting the heart,

---

Adapted from the author's Grand Rounds presentation at The Mount Sinai Medical Center on October 19, 1993. From the Division of Endocrinology and Metabolism, Johns Hopkins University, Baltimore, MD. Address reprint requests to Paul W. Ladenson, M.D., Division of Endocrinology and Metabolism, Johns Hopkins Hospital, 600 North Wolfe Street, Blalock 904, Baltimore, MD 21287-4904.



1980 March 1982

October 1982 1983

Fig. 1 Serial photographs of a patient 10 years (1980), 8 years (March 1982, October 1982), and 7 years (1983) before presentation with biventricular heart failure and myxedema.

pericardial effusions. Although echocardiographically detectable pericardial fluid is often present in hypothyroidism, affecting 30% to 80% of patients, it is rarely hemodynamically significant; there are only a two dozen cases of pericardial tamponade with hypothyroidism in the literature. The main clinical significance of pericardial effusions in this circumstance is that they can lead to misdiagnosis of heart failure. If you can imagine a patient who is fatigued, who may be experiencing dyspnea on exertion, who has edema, and has these chest x-ray findings, it is easy to see how he might be thought inadvertently to have heart failure. This mistake has two risks. First, hypothyroidism may not be recognized and treated; and second, patients may be treated with a digitalis glycoside, to which they have greater risk of developing toxicity because of the impaired metabolic clearance of digoxin in hypothyroidism. Pericardial effusions due to hypothyroidism usually resolve over a couple of months, sometimes taking as long as a year (3).

Hemodynamic Changes in Hypothyroidism

There are clinically important hemodynamic aspects of hypothyroidism and its reversal. All of us recognize that bradycardia is one of the hallmarks of hypothyroidism. In addition to this negative chronotropic effect of thyroid hormone deprivation, there is a reversible impairment of myocardial contractility. A third effect which we do not usually consider is the marked increase in peripheral vascular resistance that occurs in hypothyroidism. Clinically, the cool pale skin and diastolic hypertension that are typical of patients with hypothyroidism are manifestations of this intense peripheral vasoconstriction, which is also reversible with thyroid hormone therapy.

Reversibility by Thyroid Hormone

There is a well established myocardial dose-response curve in hypothyroid patients treated with increasing doses of thyroxine. In a study by Crowley et al. (4), for example, the ratio of the pre-ejection period, which gets longer with decreased contractility, and the left ventricular ejection time, which becomes shorter with less complete ventricular emptying, was higher before treatment in hypothyroidism. As the patients in this study were treated with incremental increases in thyroxine dose, this noninvasive parameter indicative of myocardial contractility improved progressively. In fact, enhancement of contractility can occur relatively quickly after institution of thyroid hormone replacement. In the early 1980s, we showed that when intravenous thyroxine (T4) or triiodothyronine (T3) was given to patients with hypothyroidism, there could be marked prompt enhancement of myocardial contractility, ranging from 25% to 100% complete normalization, within the first week of treatment (5,6).

Obviously this was the kind of hemodynamic response that we were hoping for in our patient with hypothyroidism. With the institution of thyroxine therapy, you can see that there was a prompt increase in his serum T4, and a steep decline in the serum TSH level (**Fig. 3**). I felt emboldened to start with full-dose L-thyroxine replacement, 150 μg/day, in this young man, who had no risk factors for coronary artery disease: no

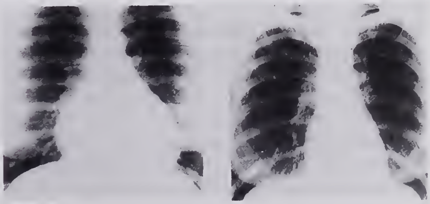


Fig. 2 Anteroposterior chest x-rays in a patient with hypothyroidism, before and after restoration of euthyroidism with thyroxine.

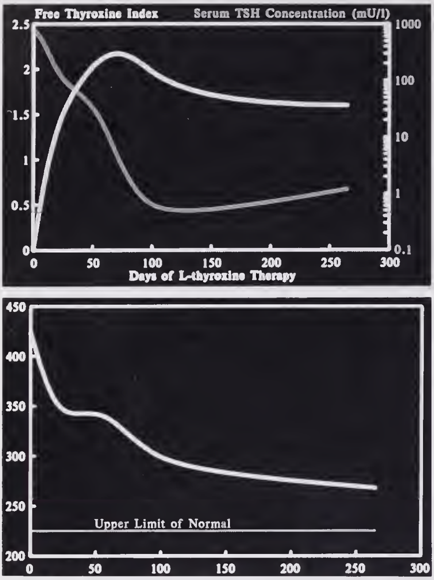


Fig. 3 Serial serum-free thyroxine index (normal, 1.25–4.25) and TSH (normal, 0.5–4.5 mU/L) concentrations (upper panel) and pulse-wave arrival time (QKd) determinations (lower panel) in a patient with hypothyroidism and heart failure receiving 150-μg thyroxine per day.

family history, a serum cholesterol that had remained remarkably within the normal range despite hypothyroidism, and no history of smoking.

I want to point out a phenomenon that is important to consider when we treat patients with hypothyroidism. Despite the fact that we left this patient on the same dose of thyroxine during this entire period, you can see that, in the first few months after treatment, his free thyroxine index was higher than it subsequently was after 6 months of therapy. This is because a patient's thyroid status determines the rate of metabolism of thyroxine itself; in the hypothyroid state, one metabolizes thyroxine less rapidly. A few important clinical implications of this fact are: in younger patients, we should not be concerned about high serum free T4 levels early in the course of treatment; but in elderly patients, one would rather not bathe tissues like the heart with excessive thyroid hormone levels early on in therapy. This is another argument for a stepwise approach to thyroxine treatment in the elderly, a point to which I will return.

Noninvasive assessments of this young man's myocardial contractility steadily improved during the first 9 months of thyroxine therapy. For the first 2 weeks, he was observed in our Clinical Research Center; subsequently, he was readmitted periodically for evaluations. Throughout, there was a progressive decline in his pulse-wave arrival time (QKd) (Fig. 3). Remembering that the upper limit of normal for the QKd is 225 milliseconds, this young man's pretreatment value was longer than any I had ever seen, 424 milliseconds, with progressive improvement. When he returned after more than 2 years of treatment, we found his QKd to be 236 milliseconds. Although this still was not completely normal, it obviously was greatly improved. This improvement was also reflected in other hemodynamic measures: an increase in his cardiac index to 2.7 L · min/m^2 and a decrease in his echocardiographically assessed end-diastolic and end-systolic ventricular dimensions. All of this was accompanied by a spectacular clinical improvement. A young man who was facing cardiac transplantation was, after 6 months of treatment, mowing lawns and delivering pizzas at night. (I can imagine few jobs as stressful as his second one.)

All of this was interesting clinically and physiologically, but not really very novel. However, this remarkable patient gave us a unique opportunity to investigate some of the molecular mechanisms responsible for the thyroid hormone's influence on the heart. An endomyocardial biopsy was obtained before thyroxine treatment because of clinical skepticism that hypothyroidism alone could explain such profound heart failure; and that biopsy showed no evidence of active myocarditis, hemosiderin deposition, or other findings for a specific form of cardiomyopathy. A second biopsy was obtained after 9 months of thyroxine treatment because of the incomplete normalization of his cardiac function, and with permission of our human studies committee. Thus, we had the extraordinary chance to use two tiny pieces of myocardium as a window on some of the molecular events that had been associated with and were potentially responsible for the improvement in his heart's contractile function.

Molecular Basis of Myocardial Dysfunction

I want to talk about the underlying physiology of how thyroid hormones alter myocardial function. During the past four decades, there have been a number of theories. The first was the rather vague suggestion that thyroxine simply increased myocardial mass. We have clearly moved beyond that to understand some of the specific molecular events that are triggered by thyroid hormones. In the 1970s, Williams and Lefkowitz (7) showed that the number of β-adrenergic receptors on rat myocardiocytes was regulated by thyroid hormone. This fit well with the notion that thyroid hormones enhanced sensitivity to catecholamine action. This postulate, I think, remains interesting, but is incompletely substantiated. There is no direct evidence that T3 regulates β-adrenergic receptor gene expression, and there are alternate explanations for the changes in β-adrenergic receptor number in altered thyroid states.

We now know that thyroid hormones affect sarcoplasmic reticulum calcium exchange. I will say more about this in a second, but the story I want to tell you is how thyroid hormones alter myosin ATPase activity. First, just an update on how thyroid hormones are believed to work at the molecular level—T4 and T3 enter cells in target tissues, from the cytoplasm they enter cell nuclei by facilitated diffusion, and then predominantly T3 binds to a nuclear receptor that has been characterized in a number of distinct isoforms and species. The T3 receptor has three important domains: one binds T3 with high affinity; a second binds to specific oligonucleotide sequences, called a T3 regulatory element, that are in the promoter regions of specific genes; and a third facilitates the binding of these receptors to specific accessory proteins and one another to form dimers, which is important for them to be active. When T3 and its receptor interact with a positive regulatory element, there is enhanced expression of that particular messenger RNA, which when translated into proteins affects the actions of thyroid hormone. We were interested in defining whether certain mRNAs implicated in the myocardial actions of thyroid hormone would be expressed in greater abundance when a patient like ours, who had been deprived of thyroid hormone, had it replaced therapeutically.

The myosin molecule that interdigitates with actin in myofibril filaments has so-called heavy chain heads that sequentially cross-link to the actin molecule to effect contraction. These molecules also possess ATPase activity, which provides the energy for these sequential cross-linking events. There are several distinct species of myosin heavy chains. α-Myosin, a homodimer comprising two α chains, has greater activity in terms of ATP hydrolysis and the speed with which it successively cross-links with actin in contraction. On the other hand, β-myosin, which comprises two myosin β chains, has lesser ATP hydrolytic and contractile properties.

You can see what direction this story is heading in in terms of thyroid hormone action. Both of the cardiac myosin heavy chain genes are encoded by genes on the short arm of chromosome 14. For more than 20 years, it has been known that, in small rodents, thyroid hormone produces a shift in the predominant isoform of myosin heavy chain being expressed. In the absence of thyroid hormone, the expression of the slower acting β-myosin heavy chain increases; whereas, in the presence of thyroid hormone, there is a shift toward α-myosin heavy chain expression, a phenomenon that had been observed in the rat and rabbit, but never previously in man.

We used the technique of reverse transcription-polymerase chain reaction, which enables one to use exceedingly small quantities of messenger RNA and to amplify to determine semiquantitatively the levels of expression of specific genes in a target tissue. For our studies, we investigated the levels of mRNA encoding α- and β-myosin heavy chains, introducing an internal control RNA sequence to make this a more quantitative observation. We found that before thyroid hormone therapy there was a relatively low level of α-myosin heavy chain gene expression in the myocardial specimen from our patient, but that after 9 months of thyroid hormone therapy, α-expression was increased 11-fold. Unlike previous animal experiments, there was little change in β-myosin heavy chain gene expression. I would also like to mention that the mRNA for atrial natriuretic peptide, which is expressed in the adult ventricle only in the setting of heart failure, was present before treatment, but no longer detectable after reversal of heart failure in our patient, a molecular marker of his myocardial recovery.

Although I have focused on thyroid hormones' effect on myosin heavy chain gene expression, and this can alter systolic function, there is a similar story related to the sarcoplasmic reticulum calcium ATPase. It is clear that T3 regulation of the SR calcium ATPase gene is another important factor affecting cardiac function, more specifically the diastolic function changes that are related to thyroid status.

Returning to the clinical realm, please remember that our patient is a very unusual one. In most hypothyroid patients with heart failure, there is intrinsic heart failure that has simply been worsened by superimposed thyroid hormone deficiency: its further impairment of contractile function and its increase in peripheral vascular resistance. These cases should focus primarily on identifying and treating that intrinsic heart disease. In addition, in the short term, one can increase contractility by administering digitalis glycosides, which were never used in our patient, and, of course, by administering the "right stuff," thyroid hormone.

Thyroxine Therapy and Ischemic Heart Disease

As you are all well aware, the kind of enhancement in myocardial contractility produced by thyroid hormone can also have deleterious consequences. When you have a patient with ischemic heart disease, decision-making about instituting thyroid hormone therapy is more difficult than it was in the earlier case.

Why should we be especially worried about atherosclerotic coronary artery disease in patients with hypothyroidism? The disorder itself can increase certain risk factors for atherosclerosis. Patients with hypothyroidism can have elevated low-density lipoprotein (LDL)-cholesterol levels because thyroid hormone regulates hepatic LDL receptor expression and, consequently, LDL clearance. Furthermore, an increase in diastolic blood pressure is typical in hypothyroidism, as I mentioned earlier. In addition, some have argued, not too convincingly, that hypothyroidism, or even autoimmune thyroid disease itself, may be independent risk factors for atherosclerosis. In any case, you can imagine what is going to happen to myocardial oxygen demands in a hypothyroid patient treated with thyroid hormone. There is going to be an increase in both heart rate and myocardial contractility, and these positive chronotropic and inotropic actions of thyroid hormone are going to increase myocardial oxygen demand; and in a patient with critically narrowed coronary arteries, coronary flow may be unable to meet this higher demand. On the other hand, all of the effects of thyroid hormone therapy are not bad in this setting. If you think about how a dilated hypothyroid heart decreases in size with thyroid hormone therapy, and remember the law of Laplace that relates wall tension to the radius of a sphere, you realize that preload can decrease with treatment of hypothyroidism. Afterload will also be reduced with the lowering of peripheral vascular resistance. Consequently, it is hard to predict in a given patient exactly what is going to happen clinically when you treat hypothyroidism in the presence of coronary disease. In fact, considerable variability in response has been observed in clinical trials. In one study from the Mayo Clinic in 1960, the results of thyroid hormone therapy in 55 patients with preexisting angina was described (8); 46 patients actually improved or stayed the same, whereas only 9, approximately 20%, had worsened angina. In fact, this was probably not different from the natural history of coronary disease in the 1950s before availability of β-blockers and long-acting nitrates.

The approach that the physicians from Mayo Clinic used is still the one that most of us recommend today: starting low and going slow, beginning L-thyroxine perhaps 25 μg or even 12.5 μg/day as a starting dose; and then increasing the dosage in 12.5- to 25-μg increments at 4- to 6-week intervals, remembering that the half-life of thyroxine is a bit longer than even 7 days in hypothyroid patients. Along the way, the patients must be monitored closely, both clinically and electrocardiographically. If ischemia worsens with thyroid hormone therapy, consider bolstering the antianginal regimen; β-blockers are particularly useful in this setting. One can occasionally compromise on the L-thyroxine dose, but my own view is that you should not compromise much; this leaves the patient not only with a low T4 and a persistently elevated TSH, but also with risk factors—higher LDL cholesterol and diastolic blood pressure—that may accelerate their coronary disease.

Coronary Surgery and Angioplasty in the Hypothyroid Patient

In general, I think that these patients ought to receive coronary surgery or angioplasty, just like euthyroid patients, based on conventional criteria. The safety of this approach has been studied in three retrospective series, two from the Mayo Clinic (9, 10) and one that we performed at the MGH (11); in none of these did hypothyroid patients undergoing coronary artery bypass graft surgery have greater mortality or duration of hospitalization than that observed in euthyroid age- and sex-matched controls. This last study did demonstrate, however, that there are certain potential complications that need to watched for in a hypothyroid patient undergoing cardiac surgery (**Fig. 4**). Intraoperatively and for the first few days after surgery, there is greater risk of heart failure. Compared with controls, patients with hypothyroidism were more likely to have postoperative gastrointestinal and neuropsychiatric problems, constipation and ileus, and neurologic problems, confusion in the night or even seizures. All of these are potentially serious, but they are manageable if recognized. Finally, patients with hypothyroidism do not have elevated fevers as well as euthyroid individuals. Despite comparable prevalences of infection in the two groups, the hypothyroid patients were only half as likely to

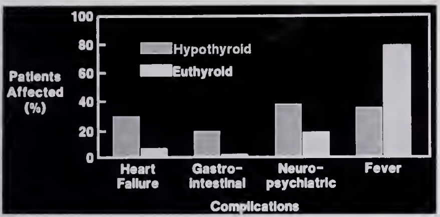


Fig. 4 Perioperative complications occurring more frequently in patients with hypothyroidism vs age- and sex-matched controls. (From Ladenson et al. Am J Med 1984; 77: 261–266).

have postoperative fever. Therefore, clinicians need to have a high index of suspicion of infection in the hypothyroid patient facing surgery. So, in general, I recommend withholding thyroid hormone therapy in patients with unstable angina, unless there is a severe complication of myxedema, like hypercapnic ventilatory failure. Then based on usual criteria, one can proceed with surgery or angioplasty, keeping in mind these potential complications.

Angioplasty is a particularly attractive option in this setting. Dr. Steven Sherman and I studied the Hopkins experience with hypothyroid patients who had angioplasties (12), again comparing them with matched controls. Their outcomes, in terms of dysrhythmias, heart failure, success of the procedure, and reocclusion were no different; the only difference, and even this was not statistically significant, seemed to be a higher rate of hematoma formation and postprocedural bleeding in the hypothyroid patients. This fits well with what we know about a von Willebrand-like defect in platelet function in hypothyroidism.

Cardiac Complications Of Hyperthyroidism

Hemodynamic Changes with Hyperthyroidism

In many senses, hyperthyroidism results in the opposite of what hypothyroidism does to the cardiovascular system. Of course, tachycardia is a clinical hallmark of thyrotoxicosis. Enhancement of contractility can be inferred from the subjective sense of palpitations and the increased systolic blood pressure that occurs with hyperthyroidism. The flushed warm skin of hyperthyroid patients is evidence of the substantial decrease in peripheral vascular resistance that occurs.

There are three cardiovascular complications that occur in thyrotoxic patients. The first is atrial tachyarrhythmias: most commonly, of course, sinus tachycardia, then atrial fibrillation, less commonly paroxysmal atrial tachycardia, and rarely atrial flutter. Mitral valve dysfunction and heart failure can also occur in hyperthyroidism.

Thyrotoxic Atrial Fibrillation

Several years ago, the nation's attention was drawn to thyrotoxic atrial fibrillation (AF), when the President developed his clinically important cardiac complication of hyperthyroidism. I do not need to tell an audience of internists that loss of the atrial kick with rapid ventricular response times and little time for diastolic filling can lead to heart failure; or that thromboembolism is a potentially irreversible and catastrophic complication of thyrotoxic atrial fibrillation. Earlier retrospective studies have reported 10% to 40% incidences of thromboembolism in thyrotoxic AF (13). Although I think these values probably reflect some ascertainment bias, it is obvious nonetheless that a minority of patients with thyrotoxic AF can suffer thromboembolism.

One of the interesting things about thyrotoxic AF is how mild hyperthyroidism can be and still cause the arrhythmia. It is well established that AF can occur when the sole chemical manifestation of hyperthyroidism is a suppressed TSH. To some extent, we can predict who is at risk of this complication. As patients grow older, AF becomes an increasingly common expression of hyperthyroidism; and males are more likely to develop this complication at any age (14). In a cohort of patients like our former president, fully 40% of thyrotoxic men in their 60s develop AF as part of thyrotoxicosis. Left atrial size is another factor that has been recognized as predisposing to development of AF (15).

Several important therapeutic issues should be addressed in the thyrotoxic AF patient. The first is anticoagulation, which I generally recommend for patients older than 50 years old, those who have histories of previous emboli, heart failure, or hypertension, or with echocardiographic evidence of left atrial enlargement. Remember that in hyperthyroidism, catabolism of the vitamin K-dependent clotting factors accelerates. So, even though their loading dose of coumarin may be the same, patients may require a lower maintenance dose. Rapid ventricular response rates can cause trouble in thyrotoxic AF patients and may be hard to control. The high doses of β-blockers, calcium channel blockers, and digoxin required can sometimes cause toxicity. In general, β-blockers are the preferred agent of ventricular rate control in this setting. They are able to slow the ventricular rate more per decrement in contractility than can calcium channel blockers. In patients with heart failure, in whom you want to control heart rate, but are worried about impairing contractility, the use of a short-term β-blocker like esmolol intravenously under close observation is warranted.

Finally, when should you cardiovert a patient with thyrotoxic AF? Useful guidelines were developed about a decade ago in an important study from Japan in which more than 150 thyrotoxic AF patients were reported (16). The authors observed, to no one's surprise, that cardioversion should generally be deferred until the patient is euthyroid. Although cardioversion may succeed initially, patients often flip back into AF if they are still hyperthyroid. Once euthyroidism is restored, all patients who reverted to sinus rhythm spontaneously did so within 4 months of being euthyroid. If you have waited that long without results, then pharmacologic or electrical cardioversion certainly is reasonable. These investigators also found that patients who had been in AF for more than 1 year and those who were older were also more likely to need intervention in the long run, probably reflecting the coexistence of intrinsic heart disease in these thyrotoxic AF patients.

Mitral Valve Dysfunction in Thyrotoxicosis

There are a few ways that hyperthyroidism can be associated with mitral valve dysfunction. One is in thyrotoxic children with biventricular failure and dilated ventricles with mitral valve leaflets that simply can no longer be apposed. This is generally reversible with the treatment of thyrotoxicosis. Another interesting way in which mitral valve dysfunction can occur in hyperthyroid patients is the well-documented association of mitral valve prolapse with autoimmune thyroid diseases (17, 18).

Hyperthyroid Heart Failure

Finally, hyperthyroid heart failure has long been a subject of controversy. Some argue that this entity just does not exist, that every patient with hyperthyroidism and heart failure really has some form or another of intrinsic heart disease. I think this is unlikely because I, like most endocrinologists and cardiologists, have seen a handful of cases of young people with no documented intrinsic heart disease who have had quite complete recoveries from heart failure with treatment only of their thyrotoxicosis. Although heart failure in these patients is in some sense high output, it is more complex than just peripheral tissue demands asking the heart to work harder than it is able. The arteriovenous oxygen tension difference in hyperthyroidism is actually reduced. In experimental animals, a form of myocarditis has been observed with experimental hyperthyroidism, but this does not appear to be the case in the few humans whose endomyocardial biopsies have been reported. The possibility that hyperthyroid heart failure represents some kind of nutritional cardiomyopathy has been raised, but never firmly established.

Hyperthyroid heart failure remains an enigmatic entity and, fortunately, one that is rare. It is challenging to treat. Emphasis needs to be on treating the thyrotoxicosis at several levels: using β-blockers carefully if ventricular rate control is an issue, administering iodide to block the release of preformed hormone from the gland, and a using drug like ipodate that blocks extrathyroidal T4-to-T3 conversion. Antithyroid medication, such as propylthiouracil, which is the drug of choice in this setting because of its impairment of T4-to-T3 conversion, is also important. Once euthyroidism is restored or approached, definitive treatment, generally with radioactive iodine, is required to be sure that hyperthyroidism will not return. I mentioned earlier the role of β-blockers and digoxin in lowering heart rate. Here the goal ought not to be decreasing the ventricular response to 80 or 90; unlike euthyroid patients, one should usually be satisfied with a ventricular response rate of 110 to 120. In the presence of peripheral vasodilatation, diuretics and vasodilators should be used with great care in the patient with hyperthyroid heart failure.

Conclusions

We have discussed the cardiovascular complications of thyroid disease, the common but innocuous pericardial effusions that occur in hypothyroidism, and the interesting heart failure that usually is a reflection of the effect of thyroid hormone deficiency on top of intrinsic heart disease, but which rarely can occur as an isolated consequence of the lack of thyroid hormone. The greatest challenge for internists in this area is the patient who has both hypothyroidism and ischemic heart disease; most of these patients can be managed well if you remember the considerations that we discussed. The most important cardiovascular consequence of hyperthyroidism is atrial fibrillation and its associated risk of thromboembolism.

Questions and Answers

Terry Davies: Let me start by saying that because cardiac function did not return totally to normal in the case you presented of the hypothyroid heart failure, does that not actually put you in my camp, which would be that you do not get heart disease unless you have heart disease?

Ladenson: Well, Terry is right that our patient has not returned completely to normal. His last ejection fraction was 37%, which is not too bad. My own theory about this is that his heart has been stretched by this long period of profound hypothyroidism and that the three-dimensional anatomy and mechanical properties of his heart have been irreversibly altered to some extent. But I certainly would agree with the general thrust of what you are saying, that this is a unusual case, and the clinician's response to a hypothyroid patient with heart failure should always be to seek the likely presence of another underlying cause.

Leslie Kuhn: Regarding the high incidence of atrial fibrillation in older patients with hyperthyroidism, what is your explanation? Is that just longer duration of hyperthyroidism, or is there underlying coronary disease? What happens to the atrial size when they are treated?

Ladenson: Those are all good questions, and I do not think that I have the answers to any of them. We know of the clinical and epidemiologic associations of higher AF risk with older age, with being a male, and with large left atrial size. My suspicion is that there is a subtle change in the conducting system with aging that predisposes people to this expression of hyperthyroidism, but the exact electrical events are not understood completely. Regarding the reversibility of left atrial enlargement, I do not recall that it was partially reported in the study that I mentioned. I think you are expressing some skepticism about which was the horse and which was the cart here: Did long-standing AF cause atrial enlargement or the other way around? I would like to reemphasize how important it is to detect subclinical hyperthyroidism in older patients with AF. Your first response in this setting should be to find a TSH in a highly sensitive assay, that is one with a detection limit of 0.1 mU/L or less. Even if the free T4 is normal, you have not ruled out very mild thyrotoxicosis as the cause of an older patient's AF.

References

1. Ladenson PW, Sherman SI, Baughman KL, Ray PE, Feldman AM. Reversible alterations in myocardial gene expression in a young man with dilated cardiomyopathy and hypothyroidism. Proc Natl Acad Sci USA 1992; 89: 5251–5255.
2. Zondek H. Das myxodemherz. Munch Med Wochenschr. 1918; 65:1180–1183.
3. Khaleeli AA, Memon N. Factors affecting resolution of pericardial effusions in primary hypothyroidism: a clinical, biochemical, and echocardiographic study. Postgrad Med J. 1982; 58:473–476.
4. Crowley WF, Ridgway EC, Bough EW, et al. Non-invasive evaluation of cardiac function in hypothyroidism: response to gradual thyroxine replacement. N Engl J Med. 1977; 296:1–6.
5. Ladenson PW, Ridgway EC. Early peripheral responses to intravenous thyroxine in primary hypothyroidism. Am J Med. 1982; 73:467–474.
6. Ladenson PW, Goldenheim PD, Ridgway EC. Rapid pituitary and peripheral tissue responses to intravenous L-triiodothyronine in hypothyroidism. J Clin Endocrinol Metab. 1983; 56:1252–1259.
7. Williams LT, Lefkowitz RJ. Thyroid hormone regulation of β-adrenergic receptor number. J Biol Chem. 1977; 252:2787–2789.
8. Keating FR Jr, Parkin TW, Selby JB, Dickinson LS. Treatment of heart diseases associated with myxedema. Prog Cardiovasc Dis. 1960; 3:364–381.
9. Hay ID. Thyroxine therapy in hypothyroid patients undergoing coronary revascularization: a retrospective analysis. Ann Intern Med. 1981; 95:456–462.
10. Weinberg AD, Brennan MD, Gorman CA, Marsh HM, O'Fallen WM. Outcome of anesthesia and surgery in hypothyroid patients. Arch Int Med 1983; 143:893–897.
11. Ladenson PW, Levin AA, Ridgway EC, Daniels GH. Complications of surgery in hypothyroid patients. Am J Med. 1984; 77:261–266.
12. Sherman SI, Ladenson PW. Percutaneous transluminal angioplasty in hypothyroidism. Am J Med. 1991; 90: 367–370.
13. Bar-sela S, Ehrenfeld M, Eliakim M. Arterial embolism in thyrotoxicosis with atrial fibrillation. Arch Intern Med. 1981; 141:1191–1992.
14. Forfar JC, Caldwell GC. Hyperthyroid heart disease. Clin Endocrinol Metab. 1985; 14:491–508.
15. Iwasaki T, Naka M, Hiramatsu K, et al. Echocardiographic studies on the relationship between atrial fibrillation and atrial enlargement in patients with hyperthyroidism. Cardiology. 1989; 76:10–17.
16. Nakazawa HK, Sakurai K, Hamada N, Momotani N, Ito K. Management of atrial fibrillation in the post-thyrotoxic state. Am J Med. 1982; 72:903–906.
17. Channick BJ, Adlin EV, Marks AD, et al. Hyperthyroidism and mitral-valve prolapse. N Engl J Med 1981; 305:497–500.
18. Marks AD, Channick BJ, Adlin EV, Kesler RK, Braitman L, Denenberg BS. Chronic thyroiditis and mitral valve prolapse. Ann Intern Med. 1985; 102:479–483.

Grand Rounds

Irritable Bowel Syndrome

LAWRENCE S. FRIEDMAN, M.D., AND RICHARD B. LYNN, M.D.

The Abdominal Man or Woman

A CASE POPULARIZED BY W. Grant Thompson who adapted it from Hutchinson is that of the Abdominal Man or Woman (1). This is a patient who usually shows up late on a Friday afternoon before a vacation.

> Incessant demand for sympathy or understanding makes the Abdominal Man or Woman a veritable vampire, sucking the vitality out of all who come near. Half an hour with him or her reduces the doctor to the consistency of chewed string and is more exhausting than all the rest of the doctor's daily visits put together. For the Abdominal Man or Woman is always discovering fresh symptoms, will not admit to any improvement, and has an objection to everything that is proposed.

This is everyone's vision (or fear) of irritable bowel syndrome in the extreme, what we call the "irritable patient syndrome" or sometimes "chronic relapsing abdominal pain." In fact, this is the exceptional case. Most patients with irritable bowel syndrome are not this extreme, and whether this example actually represents a variant of irritable bowel syndrome or not is still debated.

Diagnosing Irritable Bowel Syndrome

What is irritable bowel syndrome? First, it is chronic. The diagnosis should not be considered unless symptoms have been present continuously or at least intermittently for 3 months or more. Second, it is functional. No specific structural, biochemical, or infective etiology has been identified. Third, it belongs to a family of functional gastrointestinal disorders that share features of irritable bowel syndrome, both in their clinical manifestations and in the fact that these occur together: chest pain of unexplained origin, nonulcer dyspepsia, and biliary dyskinesia.

A working definition of irritable bowel syndrome, well known to most clinicians, is that of abdominal pain associated with altered bowel habits: constipation, diarrhea, or classically, alternating constipation and diarrhea (2). It has long been debated whether painless diarrhea is a variant of irritable bowel syndrome. Most of us consider it to be, although for purposes of research, abdominal pain is usually required in the case definition. On the other hand, constipation alone, without pain, is usually not considered part of the spectrum of irritable bowel syndrome.

In 1978, Manning and his colleagues (3) identified in patients with gastrointestinal symptoms six criteria that helped to distinguish patients with irritable bowel syndrome from patients with other gastrointestinal diseases. These have come to be called the Manning criteria and they include:

- abdominal distention
- relief of pain after a bowel movement
- more frequent stools with the onset of pain
- looser stools with the onset of pain
- passage of mucus
- a sensation of incomplete evacuation

It is quite helpful to query the patient about these symptoms in making a diagnosis of irritable bowel syndrome. In a 21-year-old woman with gastrointestinal symptoms suggestive of irritable bowel syndrome, the presence of only two of these six criteria yields a 64% likelihood that, in fact, the patient has irritable bowel syndrome. When all six criteria are present, 92% of such patients

---

Adapted from the Grand Rounds presentation of Dr. Friedman at the Department of Medicine, Mount Sinai School of Medicine, New York, NY, on April 18, 1995. From the Department of Medicine, Harvard Medical School, and the Gastrointestinal Unit, Massachusetts General Hospital, Boston, Massachusetts and the Division of Gastroenterology and Hepatology, Jefferson Medical College, Philadelphia, Pennsylvania.

have irritable bowel syndrome (4). As patients age, the diagnostic sensitivity of these criteria decrease.

A recent consensus conference in Rome incorporated the Manning criteria into a case definition of irritable bowel syndrome, often referred to as the Rome criteria for irritable bowel syndrome (5). The definition is as follows: (a) abdominal pain, or discomfort relieved with defecation or associated with a change in the frequency or consistency of stool; and (b) three or more of the Manning criteria: altered stool frequency, altered stool form, altered stool passage, the passage of mucus, and bloating or abdominal distention.

Patients with irritable bowel syndrome often have symptoms that are not necessarily related to the colon, where most early research in irritable bowel syndrome focused (6). These symptoms include: gastroesophageal reflux with heartburn, dysphagia, globus sensation (now thought to be a manifestation of gastroesophageal reflux), chest pain of undetermined etiology, urologic dysfunction, gynecologic complaints, and fatigue. Thus, irritable bowel syndrome has overtones of a systemic disease, even if one cannot identify specific pathophysiologic abnormalities.

Epidemiology

Irritable bowel syndrome is a remarkably common disease. Surveys have shown that between 10% and 22% of the population fulfill the criteria for irritable bowel syndrome (7). In fact, gastrointestinal complaints show up on standardized health questionnaires in about 50% of individuals. The prevalence of irritable bowel syndrome is similar in younger and older patients. The onset is generally in young adulthood, and the diagnosis is generally secure once made. The prevalence of irritable bowel syndrome is constant in age cohorts from young adulthood on, although some people lose the criteria whereas others enter the arena of irritable bowel syndrome, so the population is not necessarily stable, even though the prevalence is. Of patients with irritable bowel syndrome, 40 to 50% actually seek medical attention. This subset of patients may differ from the population of patients with irritable bowel syndrome that do not consult a physician (see below). In this country women predominate; about two thirds of the patients who consult a physician for the disorder are women. Interestingly, in community survey studies in the United States, little over half of the patients are women, whereas in India the majority of patients consulting a physician are men. It is not clear that the disorder has a sex predilection; rather gender may affect who consults a physician. Cultural and psychosocial factors may account for the reported sex differences.

Pathogenesis

The literature is remarkably confusing regarding the pathogenesis of irritable bowel syndrome. The current thinking is that the disorder is really a heterogeneous group of diseases and that multiple abnormalities may account for the syndrome, but in a given patient only one or another may be present. Four categories of pathophysiologic abnormalities have been described. Abnormalities have been reported in colonic motility, in which most of the early research focused, small bowel motility, and visceral sensation. In addition, study of the role of psychosocial factors has provided insight into the pathophysiology of the disorder.

In recordings of the electrical and motor activity of the colon, periodic slow waves are found, and, from time to time, spike potentials are superimposed on these slow waves. Only with spike potentials do muscular contractions of the colon occur. The contractions may be segmental or peristaltic depending on the type of spike potential. Early studies of the pathogenesis of irritable bowel syndrome identified abnormalities of the colonic slow wave pattern. In the healthy adult the frequency of colonic slow waves averages 6 cpm. Studies by Snape and his colleagues (8) suggested that patients with irritable bowel syndrome had an increased frequency of 3-cpm slow-wave activity, and for many years this 3-cpm pattern was considered a marker of irritable bowel syndrome. However, other investigators have not been able to confirm this finding. Even in patients with so-called constipation-predominant irritable bowel syndrome, abnormalities of colonic slow waves have not been identified consistently, and their significance remains uncertain.

Other investigators have looked at the effects of various stimuli on the colon and found reproducible abnormalities in response to the intake of food and to certain stresses, suggesting abnormal excitability of the colon (9). A problem with these studies is that they look only at acute stresses, whereas most patients with irritable bowel syndrome have symptoms not in response to acute stresses, but rather spontaneously, in response perhaps to chronic life stresses. As yet, the methodology to study colonic motility on a long-term basis has been limited. Some investigators reported a specific abnormality of cecal transit and motility in patients with diarrhea-predominant irritable bowel syndrome (10), but the significance of these findings is also uncertain.

Alterations in small bowel motility have been reported. Electrical and pressure recordings in the small bowel have led to several observations of uncertain significance. There are two patterns to motility in the small bowel: a fed pattern and a fasting pattern. Most attention has focused on the fasting pattern, in which there is a period of quiescence with no or little electrical activity (phase 1), after which the small bowel starts to "wake up" and there are periodic electrical potentials that lead to contractions of the bowel, so-called discrete clustered contractions (phase 2). Subsequently, phase 3 of the migrating motor complex occurs, in which there are high-amplitude waves that start in the distal stomach, travel through the proximal small bowel, and extend distally to "sweep out" the small bowel, resulting in the so-called "housekeeping function" of the migrating motor complex. Phase 3 contractions persist during sleep, but the phase 2 contractions disappear during sleep. Some investigators (11) have observed abnormalities of phase 2 potentials and associated contractions in patients with irritable bowel syndrome. They argued that symptoms of irritable bowel syndrome do not occur when the patient is asleep, and the absence of phase 2 electrical activity during sleep may account for this observation. However, other investigators have compared the frequency of discrete clustered contractions in phase 2 in patients with irritable bowel syndrome and controls at different points along the small bowel and found no differences (12).

The most consistent results have been obtained from studies that investigate altered visceral sensation in irritable bowel syndrome. It has long been known that patients with irritable bowel syndrome have a lower threshold for perceiving pain when a balloon or a volume of gas is inserted into the bowel, whether it is the rectum, colon, or small bowel (13). The increased sensitivity is not caused by an increase in the frequency of contractions in response to distention, nor is it caused by altered compliance of the bowel. Rather, these patients are thought to have altered visceral sensation. The question arises as to where the alteration in sensation is located: the nerves of the gut, the nerves that lead to the brain, or the brain itself? This is the subject of current investigation.

Recent observations suggest that psychosocial factors may affect the presentation of irritable bowel syndrome. A number of studies have shown that there is an increased prevalence of psychiatric diagnoses in patients who have irritable bowel syndrome, namely personality disorders, anxiety, depression, and somatization (14). Specifically, the higher frequency of these disorders is found in patients who consult a physician, not in those who do not seek medical attention. This latter group of "nonpresenters" with irritable bowel syndrome is no different psychologically from a control population. These observations have been interpreted to imply that patients with irritable bowel syndrome who seek medical attention exhibit health-care-seeking behavior (15). Put succinctly, a superimposed psychiatric disorder may convert a nonpatient with symptoms of irritable bowel syndrome into a patient. This summary is simplistic, because not all patients with irritable bowel syndrome who present to a physician have a psychiatric disorder.

Recent studies have tried to explain the role of psychosocial factors in irritable bowel syndrome. Whitehead and Crowell (16) postulated that presentation to a physician might relate to early life experience. Recent attention has focused on the role of physical or sexual abuse in these patients. A survey of residents of Olmsted County, Minnesota, has shown an association between a self-report of abuse through a validated questionnaire and the presence of gastrointestinal symptoms (17). In this population, about 40% of all women and 11% of all men reported a past history of abuse. Among patients with irritable bowel syndrome, 50% reported a history of abuse, whereas among those without irritable bowel syndrome, 23% reported a history of abuse. Similar percentages hold for patients with dyspepsia and heartburn, which may be related to irritable bowel syndrome. A history of abuse may have an effect on the patient's perception of the symptoms, tolerance for symptoms, desire to seek medical care, or all three.

Of the population of Olmsted County, the percentage of patients seeing a physician for gastrointestinal symptoms among healthy controls without a history of abuse was just over 10% (17). In this study, among healthy controls with a history of abuse the frequency was no different. By contrast, those with irritable bowel syndrome and a history of abuse had frequency of seeing a physician of more than 50%, whereas those with irritable bowel syndrome and no history of abuse had a slightly lower frequency of seeing a physician.

Investigators at the University of Alabama at Birmingham have speculated (18) that physical and sexual abuse lead to two problems that affect the way a patient relates to the health care system. One is the acute trauma that causes pain and tissue damage, which may lead to disruption of physical function centered around the perineal area. The other is an increased risk of psychiatric syndromes that may affect tolerance to pain and alter the response to stress. Previously traumatized patients may also have increased stress in their life. All these factors in combination are presumed to lead to increased complaints of physical symptoms, usually centered around chronic or recurrent pain, particularly in the pelvic or lower abdominal area. Such persons may also have altered beliefs about how to deal with pain and impaired coping strategies. The result is physical disability and health-care-seeking behavior. This model needs to be confirmed in additional studies, but it represents a reasonable explanation as to why some patients with irritable bowel syndrome, at least those with lower abdominal pain as a major component, may present to a physician, but others do not.

The role of stress is often discussed in the context of irritable bowel syndrome. Patients often believe that stress has caused their symptoms. However, we have never met a patient with any disease who did not have stress in his or her life. In fact, most studies suggest that there is no increased frequency of stress in the lives of patients with irritable bowel syndrome (19). Clearly, psychosocial factors do not cause the symptoms, although how people respond to stress appears to be an important determinant in the decision to seek attention for their problem. Any stressful event may exacerbate gastrointestinal symptoms and may precipitate the decision to see a physician.

What is the cause of irritable bowel syndrome? We do not know the answer, but we have hypotheses (20). Considering all the data summarized above, we may conclude that irritable bowel syndrome represents an overresponsiveness of the intestinal motor apparatus to stress or other stimuli, as suggested by increased electrical and motor activity of the gut. On the other hand, the abnormal electrical and motor activity could be epiphenomena of some other underlying abnormality. The other major possibility is that there is an abnormal perception of external stimuli by the central nervous system. Currently, this hypothesis is favored.

Until now, patients have generally been lumped together in studies without regard to whether they have diarrhea or constipation, or both, as their predominant symptom. It has been suggested (20) that we need to stratify patients for their predominant symptom. We also need to assess the response to standardized, reproducible, physiologic stimuli. Prolonged ambulatory recordings of electrical or motor activity in the small or large intestine over time may be particularly helpful. However, the methodology still must be standardized. Finally, in such studies we must control for confounding variables, including diet, the preparation of the bowel, stress, psychopathology, the intrinsic variability of the patient, and the methodology.

Several investigators have begun to look at animal models that ultimately may have relevance to irritable bowel syndrome. (Actually, it is difficult to imagine an animal model of irritable bowel syndrome.) What has been examined is the neurologic organization of the brain as it relates to the bowel. Specifically, using neural tracers injected into individual organs, the vagal motor and sensory pathways can be mapped. In this way it has been learned that the vagal motoneurons innervating the colon and cecum are an anatomically distinct population of neurons in the dorsal vagal complex (21). Whether abnormalities of neurotransmission in the dorsal vagal complex can be identified in patients with irritable bowel syndrome, or whether specific neurotransmitter agonists or antagonists can alter the symptoms, remains to be seen.

Approach to the Patient

The most important step in managing a patient with irritable bowel syndrome is to tell the patient what he or she has, not what he or she does not have (22). The traditional approach has been to make a diagnosis of irritable bowel syndrome by exclusion. With some degree of confidence, however, the Manning criteria can be used to tell the patient, even after the initial history and physical examination, that he or she likely has irritable bowel syndrome. There are obvious advantages to presenting the diagnosis in a positive manner to the patient.

It is also important to identify the psychosocial factors that may have precipitated the patient's presentation, because dealing with the stress may alleviate the physical complaints (23). It is not clear whether patients with irritable bowel syndrome have more pain or perceive pain more acutely than nonpresenters, but there is no question that focused psychotherapy goes a long way to alleviating their complaining, if not their actual complaints.

The spectrum of presentation is wide. According to the working definition, the symptoms may be chronic or intermittent; there may be constipation, diarrhea, or both; there may be pain and in a few cases there may be no pain; symptoms may relate specifically to meals or may not; and there may be lifestyle factors that the patient reports, whereas other patients may not report any impact on lifestyle. One has to be sensitive to the multiplicity of ways in which this syndrome can present.

The physical examination is usually unremarkable. A limited laboratory workup seems reasonable but should not be done to excess. Generally, a complete blood count, erythrocyte sedimentation rate, and chemistry panel are performed to screen for evidence of systemic disease. A flexible sigmoidoscopy is often performed. Clearly, sigmoidoscopy should be performed in anyone older than 40 years of age to screen for colorectal neoplasms. In the younger patient, routing sigmoidoscopy may not be necessary in patients with typical symptoms of irritable bowel syndrome unless there is a suspicion that the patient might have inflammatory bowel disease. If diarrhea is the predominant symptom, there ought to be a limited workup to screen for alternative causes of diarrhea, for example, parasitic infections; the stool should be examined for fecal leukocytes and, in some cases, excessive fat, which is suggestive of a malabsorption syndrome. The possibility of thyroid disease might be assessed; hyperthyroidism can cause hyperdefecation, which can mimic irritable bowel syndrome, and hypothyroidism causes constipation.

It is debated whether sigmoidoscopic biopsies are of value. They are useful in detecting variants of inflammatory bowel disease, such as lymphocytic or collagenous colitis, which occur in middle-aged women. In these disorders, colonic mucosal biopsy is diagnostic even though the endoscopic appearance of the mucosa is normal. These patients often present with a secretory type of diarrhea, a much higher volume than would be expected with irritable bowel syndrome.

There are a number of conditions that may be confused with irritable bowel syndrome. In practice, think about lactose intolerance, because it is so common and can occur in patients who also have irritable bowel syndrome. A trial of withdrawal of milk from the diet is probably the most cost-efficient way to make this diagnosis. Many patients ingest sugarless food, drinks, or gums sweetened with sorbitol or fructose, which can cause altered bowel habits, specifically diarrhea and gaseousness. Numerous medications contain sorbitol as a filler which can cause such symptoms, and some medications themselves cause gastrointestinal side effects and have to be considered (24). Patients with depression or panic disorders often present with irritable bowel-like symptoms as a predominant manifestation of the psychiatric illness. There is clearly a complex relationship between the gastrointestinal tract and the psyche.

Management

The physician-patient relationship must be based on trust and understanding. Presenting the diagnosis in a positive manner is the cornerstone of therapy. Searching exhaustively for some identifiable abnormality should be discouraged.

Initial therapy almost always focuses on dietary modification and the use of fiber. Specific therapy may depend on the dominant symptom (**Table 1**). However, one should resist the temptation to keep trying medications if symptoms do not improve immediately. An interesting review of therapeutic trials in irritable bowel syndrome has concluded that "not a single study offers convincing evidence that any therapy is effective in treating the IBS symptom complex" (25). This is rather a strong statement. However, it actually does not mean that therapy is not effective. In fact, specific drugs often are useful in the treatment of individual patients with irritable bowel syndrome. On the other hand, published studies have been flawed: they have been small, have included patients with varying symptom complexes, and have often been weak statistically. In general, it is the methodology that one can criticize rather than the lack of efficacy of the drugs. There is, in fact, a remarkable placebo response in irritable bowel syndrome. Actually, every therapy can be said to work well, because the placebo response averages 63% to 71%.

We generally rely on fiber supplementation as a first step, and for 80% of patients this may be all that is necessary. Bran or one of the synthetic fibers such as polycarbophil or psyllium, the active ingredient of Metamucil, are clearly effective in constipation-predominant irritable bowel syndrome and decrease intestinal transit time. It is debated whether fiber therapy is effective in other variants of irritable bowel syndrome. Some recent studies have suggested that as many as 55% of patients may not tolerate bran (26). This may be particularly true if bran or psyllium is introduced in a large dose and given for only a short trial period. If the dose is initially low, a teaspoon a day for example, and increased grad-

TABLE 1

Medications for Irritable Bowel Syndrome

Drug type and examples (suggested starting dose)[a]	Indications
Fiber supplements	Constipation-predominant IBS: may be tried for any IBS symptom complex
Wheat bran (½ cup to 1 bowl qd, 10–30 g qd)	
Psyllium (½ to 1 tbsp qd to bid)	
Ispaghula (20 g qd)	
Antispasmodics (anticholinergics)	Pain-predominant IBS; prevention of predictable postmeal abdominal pain
Tincture of belladonna (5–10 drops po tid)	
Dicyclomine (10–20 mg tid to qid)	
Cimetropium bromide (50 mg tid)	
Antidiarrheals	Diarrhea-predominant IBS
Diphenoxylate (2.5–5 mg qid)	
Loperamide (2 mg qd to bid)	
Cholestyramine (½–1 pck qd to bid)	
Tricyclic antidepressants	Pain-predominant IBS; chronic pain syndrome
Amitriptyline (10–25 mg qhs)	
Desipramine (50 mg qhs)	
Prokinetic agents	Constipation-predominant IBS
Cisapride (5–10 mg qhs)	
Domperidone (10–20 mg qid)	
Benzodiazepine-type anxiolytics	Anxiety, use for brief periods
Somatostatin agonists	Experimental
Calcium channel blockers	Experimental
Gonadotropin-releasing hormone analogs	Experimental
Serotonin receptor antagonists	Experimental
Kappa opioid agonists	Experimental

Adapted, with permission, from Lynn RB, Friedman LS. Irritable bowel syndrome. N Engl J Med. 1993;329:1943.

[a] Abbreviations: qd, every day; bid, twice a day; po, by mouth; tid, three times a day; qid, four times a day; qhs, at bedtime; IBS; irritable bowel syndrome.

ually up to a tablespoon of psyllium once or twice a day or a cup of bran a day, the gaseousness and discomfort that often accompanies the initiation of these agents often settles down. We find that psyllium can be used in almost all patients with irritable bowel syndrome and that many patients respond to it in combination with a healthy dose of reassurance.

Antispasmodics, usually anticholinergic agents, have some benefit in irritable bowel syndrome and are the most commonly prescribed drugs. They may be particularly helpful in pain-predominant irritable bowel syndrome, in which they can be given before the anticipated pain, such as before a meal if pain occurs predictably after the meal. Antidiarrheals such as loperamide or diphenoxylate with atropine can be used in the occasional patient who has severe diarrhea or rectal urgency that is socially compromising.

A recent meta-analysis of studies of antispasmodics in the treatment of irritable bowel syndrome has shown them to be superior to placebo in improving pain and overall well-being (27). The results of this meta-analysis are limited by the quality of the individual studies, each of which has serious methodologic flaws. Nevertheless, in an individual case the use of an antispasmodic can be helpful. Cimetropium is a newer antispasmodic that is thought to be specific for gastrointestinal smooth muscle. Probably the most commonly used antispasmodic is dicyclomine. Hyoscamine is the active ingredient in Levsin, which is also commonly used.

We occasionally use tricyclic antidepressants in patients with pain-predominant irritable bowel syndrome based on their use in other pain syndromes. In a retrospective review of the use of antidepressants in 138 patients with irritable bowel syndrome (28), the cumulative response rate was more than 80%, and complete remission was achieved in 60%. In nonresponders, there was some additional benefit to switching to another antidepressant. In some patients with diarrhea a therapeutic trial of the bile salt binding resin cholestyramine may be helpful, because in some patients bile salt malabsorption from the terminal ileum may contribute to diarrhea. Cisapride, a prokinetic agent, has been shown at least anecdotally to have some benefit in constipation-predominant irritable bowel syndrome. We try to avoid anxiolytics because of their tendency to be addicting and because they are probably of little use for long-term treatment.

Some alternative therapies have been shown to be helpful in small randomized controlled studies (29). Focused psychotherapy, performed by either the gastroenterologist or psychiatrist may be useful in otherwise refractory patients. Hypnotherapy has been shown to be beneficial in controlled studies (30), as have biofeedback or combinations of the above. Such approaches should generally be reserved for exceptional cases. A number of experimental medications have found their way into the armamentarium. There is interest in somatostatin agonists, which retard bowel motility; calcium channel blockers, which may be helpful in patients with diarrhea because the retard motility; gonadotropin-releasing hormone; serotonin receptor antagonists, which may decrease electrical sensitivity and improve the rectal urgency that may be associated with irritable bowel syndrome; and kappa opioid agonists, which decrease visceral pain. None of these has been studied in a controlled fashion. Mathias and his colleagues (31) have treated more than 200 patients with "functional bowel disease" with the gonadotropin-releasing hormone agonist leuprolide for up to 5 years. In these patients, symptomatology was reported to improve significantly in the group as a whole. These patients had "functional bowel disease," not specifically irritable bowel syndrome. Many of the responders were young women and may have had endometriosis, which enters into the differential diagnosis of irritable bowel-like symptoms. It is hard to know how to apply these results to difficult patients with irritable bowel syndrome. Controlled studies are needed.

Conclusion

Irritable bowel syndrome is a common, often distressing problem in clinical practice. A few words of wisdom are useful to keep in mind. First, if you want people to think you are wise, agree with them. This reinforces our message that a strong physician-patient bond is essential. Second, there is nothing so bad that it cannot be made worse. In other words, do not overmedicate. Finally a wise man once said, "a reliable set of bowels is worth more to a man (or woman) than any quantity of brains."

Questions and Answers

Dr. Sachar: I wonder if you would comment on a few areas that you did not have the opportunity to address. First, is there any evidence for any genetic component to the irritable bowel syndrome? Second, although you talk about the Manning criteria as the basis for a positive diagnosis, you did not really touch on the negative symptoms that would probably exclude irritable bowel syndrome. Would you discuss some of those red flags that would militate against it?

Friedman: I am not aware of any genetic component to irritable bowel syndrome, although the disorder is common and may seem to run in families. Because psychosocial factors and environmental factors play a role, it would be difficult to sort out environmental from genetic factors in an illness that is so heterogeneous.

Regarding "red-flag" symptoms, gastrointestinal bleeding, either occult or overt, always requires evaluation. Patients with systemic symptoms, such as fever or weight loss, also should be evaluated for underlying systemic disease, as should those with lymphadenopathy and hepatosplenomegaly. There may be some abdominal tenderness on examination in a patient with irritable bowel syndrome, particularly in the left lower quadrant, but there should not be any guarding or rebound tenderness. Any objective abdominal findings such as a mass should be evaluated and must not be assumed to be related to irritable bowel syndrome alone.

Question: Is there any association between irritable bowel syndrome and this subsequent development of diverticulosis or any other diseases for that matter?

Friedman: Not directly. One might suppose that colonic spasm due to an irritable bowel would predispose patients to diverticulosis, but there is no convincing evidence that it does. Nevertheless, the two disorders frequently co-exist. In fact, patients with diverticulosis who complain of left lower quadrant pain and altered bowel habits may have two diseases, that is, superimposed irritable bowel syndrome and diverticulosis. It is interesting that diverticulosis is a disease of Western society, but irritable bowel syndrome seems to be a worldwide disorder; it may not be confined to populations in which the intake of bran is low. A report from China, for example, showed that the prevalence of irritable bowel was just as high there as it is in the United States (32).

Question: What is the relationship between irritable bowel syndrome and sleep and nocturnal symptoms?

Friedman: Symptoms of irritable bowel syndrome almost never awaken the patient from sleep nor do they occur at night. This has been attributed to the absence of phase 2 activity of the migrating motor complex at night. On the other hand, there is recent evidence that patients with irritable bowel syndrome have an increased frequency of sleep disturbances, including nightmares and altered rapid-eye-movement sleep patterns. The significance of this association is not clear, but it suggests the possibility of a primary central nervous system abnormality that may affect the reticular activating system and the area of the medulla that innervates the gut.

Question: In China, when they got rid of the "iron rice bowl," did the disease become more common?

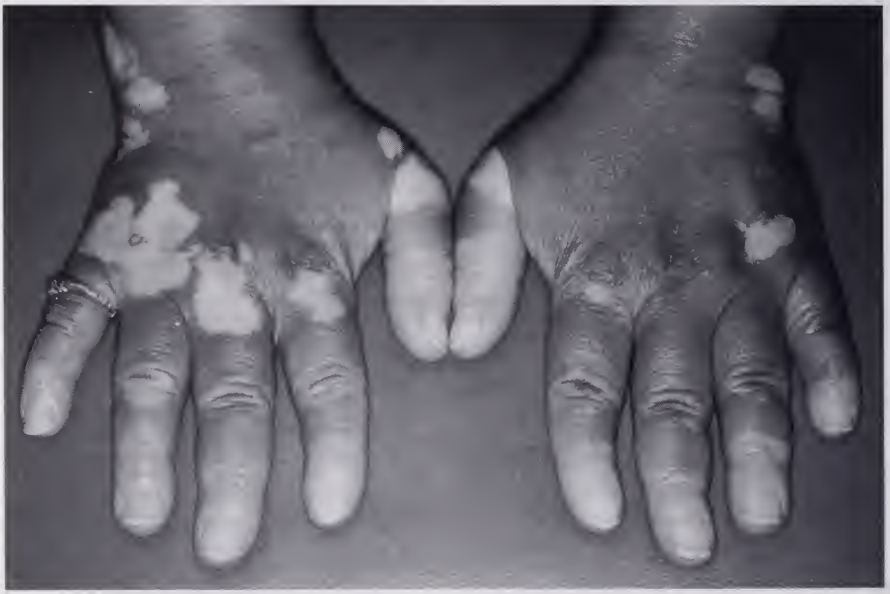
Friedman: To my knowledge, this has not been well studied.

References

1. Hutcheson R. The chronic abdomen. Br Med J. 1923;1:667.
2. Lynn RB, Friedman LS. Irritable bowel syndrome. N Engl J Med. 1993;329:1943.
3. Manning AP, Thompson WG, Heaton KW, et al. Towards positive diagnosis of the irritable bowel. Br Med J. 1978;2:653.
4. Talley NJ, Phillips SF, Melton LJ, et al. Diagnostic value of the Manning criteria in irritable bowel syndrome. Gut. 1990;31:77.
5. Thompson WG, Drossman DA, Funch-Jensen P, et al. Functional bowel disorders and functional abdominal pain. In: Drossman DA, Richter JE, Talley NJ, et al., eds. The functional gastrointestinal disorders. Boston: Little, Brown, 1994:115.
6. Smart HL, Nicholson DA, Atkinson M. Gastro-oesophageal reflux in the irritable bowel syndrome. Gut. 1986;27:1227.
7. Talley NJ, Zinsmeister AR, Van Dyke C, et al. Epidemiology of colonic symptoms and the irritable bowel syndrome. Gastroenterology. 1991;101:927.
8. Snape WJ, Carlson GM, Cohen S. Colonic myoelectric activity in the irritable bowel syndrome. Gastroenterology. 1976;70:326.
9. Welgan P, Meshkinpour H, Hoehler F. The effect of stress on colon motor and electrical activity in irritable bowel syndrome. Psychosom Med. 1985;47:139.
10. Cann PA, Read NW, Brown C, et al. Irritable bowel syndrome: Relationship of disorders in the transit of a single solid meal to symptom patterns. Gut. 1983;24:405.
11. Kellow JE, Eckersley GM, Jones M. Enteric and central contributions to intestinal dysmotility in irritable bowel syndrome. Dig Dis Sci. 1992;37:168.
12. Gorard DA, Libby GW, Farthing MJG. Ambulatory small intestinal motility in "diarrhoea" predominant irritable bowel syndrome. Gut. 1994;35:203.
13. Lasser RB, Bond JH, Levitt MD. The role of intestinal gas in functional abdominal pain. N Engl J Med. 1975;293:524.
14. Schwarz SP, Blanchard EB, Berreman CF, et al. Psychological aspects of irritable bowel syndrome: Comparisons with inflammatory bowel disease and nonpatient controls. Behav Res Ther. 1993;31:297.
15. Sandler RS, Drossman DA, Nathan HP, et al. Symptom complaints and health care seeking behavior in subjects with bowel dysfunction. Gastroenterology. 1984;87:314.
16. Whitehead WE, Crowell MD. Psychologic considerations in the irritable bowel syndrome. Gastroenterol Clin North Am. 1991;20:249.
17. Talley NJ, Fett SL, Zinmeister AR, Melton J III. Gastrointestinal tract symptoms and self-reported abuse: A population-based study. Gastroenterology 1994;107:1040.
18. Scarinci IC, McDonald-Haile J, Bradley LA, et al. Altered pain perception and psychosocial features among women with gastrointestinal disorders and history of abuse: A preliminary model. Am J Med. 1994;97:108.
19. Camilleri M, Neri M. Motility disorders and stress. Dig Dis Sci. 1989;34:1777.
20. McKee DP, Quigley EMM. Intestinal motility in irritable bowel syndrome: Is IBS a motility disorder? Part 1. Definition of IBS and colonic motility. Dig Dis Sci. 1993;38:1761.
21. Altschuler SM, Escardo J, Lynn RB, Miselis RR. The central organization of the vagus nerve innervating the colon of the rat. Gastroenterology. 1993;104:502.
22. Lynn RB, Friedman LS. Irritable bowel syndrome: Managing the patient with abdominal pain and altered bowel habits. Med Clin North Am. 1995;79:373.
23. Drossman DA. Diagnosing and treating patients with refractory functional gastrointestinal disorders. Ann Intern Med. 1995;123:688.
24. Johnston KR, Govel LA, Andritz MH. Gastrointestinal effects of sorbitol as an additive in liquid medications. Am J Med. 1994;97:185.
25. Klein KB. Controlled treatment trials in the irritable bowel syndrome: A critique. Gastroenterology. 1988;95:232.
26. Francis CY, Whorwell PJ. Bran and irritable bowel syndrome: Time for reappraisal. Lancet. 1994;344:39.
27. Poynard T, Naveau S, Mory B, Chaput JC. Meta-analysis of smooth muscle relaxants in the treatment of irritable bowel syndrome. Aliment Pharmacol Ther. 1994;8:499.
28. Clouse RE, Lustman PJ, Geisman RA, Alpers DH. Antidepressant therapy in 138 patients with irritable bowel syndrome. A five-year clinical experience. Aliment Pharmacol Ther. 1994;8:409.
29. Guthrie E, Creed F, Dawson D, et al. A controlled trial of psychological treatment for the irritable bowel syndrome. Gastroenterology. 1991;100:450.
30. Prior A, Colgan SM, Whorwell PJ. Changes in rectal sensitivity after hypnotherapy in patients with irritable bowel syndrome. Gut. 1990;31:896.
31. Mathias JR, Clench MH, Reeves-Darby VG, et al. Effect of leuprolide acetate in patients with moderate to severe functional bowel disease: Double-blind, placebo-controlled study. Dig Dis Sci. 1994;39:1155.
32. Bi-Zhen W, Qi-Ying P. Functional bowel disorders in apparently healthy Chinese people. Chin J Epidemiol. 1988;9:345.

Dermatology Notes

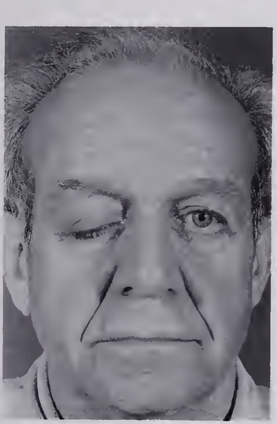
Vitiligo

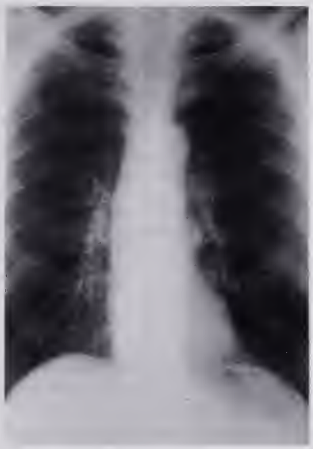


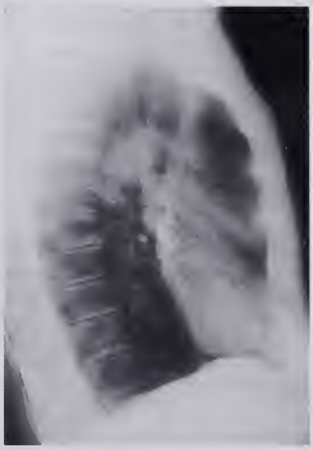
Vitiligo. Note the sharply demarcated patches of depigmentation symmetrically distributed on the hands.—*Marsha Gordon, M.D., and Mark Lebwohl, M.D.*

Ophthalmology Notes

Myasthenia Gravis







A 60-YEAR-OLD white male without significant past medical history presented to the emergency room because of abrupt onset of painless closing of his right eye. The work-up for cranial nerve palsy and cerebrovascular accident was negative; the patient was sent home. Two days later at a routine appointment at the ophthalmology clinic for refraction which had been made 6 months previously, the patient underwent neuro-ophthalmic evaluation. This examination demonstrated isolated ptosis of the right eye which resolved completely after the administration of 10 mg edrophonium IV (Tensilon Test). Thyroid function testing showed hypothyroidism, and chest radiography revealed an anterior mediastinal mass confirmed by excisional biopsy to be a thymoma. The patient was treated with oral pyridostigmine and levothyroxine and noted some improvement in his symptoms. Several months after his initial presentation he passed away at home. An autopsy was not performed.—*Arif Khan, M.D., The Mount Sinai Medical Center, New York, NY 10029.*

Gastrointestinal Histoplasmosis in a Patient with Acquired Immune Deficiency Syndrome

JAMAL RAZA, M.D., MICHAEL T. HARRIS, M.D., AND JOEL J. BAUER, M.D.

Abstract

In otherwise healthy individuals, disease caused by the fungus *Histoplasma capsulatum* manifests itself as a self-limiting pulmonary infection. Dissemination of the organisms may occur in a setting of compromised cellular immunity. Gastrointestinal involvement occurs in many such cases, but rarely it is clinically obvious and the disease seldom comes to the attention of the general surgeon. However, with the increasing incidence of acquired immunodeficiency syndromes, general surgeons are managing more of these patients than in the past. In our report, we describe a patient with acquired immunodeficiency syndrome who presented with gastrointestinal histoplasmosis.

INFECTION BY THE DIMORPHIC FUNGUS *Histoplasma capsulatum* is an endemic condition in the Ohio-Mississippi Valley, first described in 1908 (1). The organism grows in mycelial form in the environment and converts to yeast form at 37°C. The clinical presentation and morphologic lesions of histoplasmosis depend on host resistance and immunocompetence. The fungus may induce latent asymptomatic infection, self-limited primary pulmonary involvement, chronic progressive or cavitary pulmonary disease, rapid, widely disseminated systemic involvement, and localized lesions in one or more organs (2). Disseminated histoplasmosis is a rare opportunistic infection usually occurring in a setting of compromised cellular immunity with gastrointestinal involvement in many patients. Because acquired immunodeficiency syndrome (AIDS) has assumed epidemic proportions since its recognition, the general surgeon can expect to manage patients with gastrointestinal histoplasmosis, and must consider this mycotic infection in the differential diagnosis of any AIDS patient who has gastrointestinal complaints.

Although disseminated histoplasmosis has been adequately documented, the literature is limited with respect to reports of gastrointestinal histoplasmosis. We therefore report a case of disseminated histoplasmosis in a patient with AIDS with gastrointestinal involvement manifested as small bowel obstruction.

Case Report

A 25-year-old Hispanic woman, diagnosed with AIDS when she was 23-years-old, developed severe epigastric and right upper quadrant pain with nausea and vomiting for 1 week.

The patient had been admitted 8 months earlier for intermittent crampy abdominal pain, lower abdominal tenderness, watery diarrhea, and fever. At that time she was leukopenic with a normal amylase on laboratory examination. The results of plain abdominal x-rays and HIDA scan were normal. Abdominal ultrasound revealed a large echogenic liver, slightly enlarged and normal gallbladder and ducts. The patient spiked temperatures to 103°F with watery diarrhea four to five times per day, which became tinged with blood. Sigmoidoscopic findings were suggestive of colitis with yellowish plaques in the rectosigmoid area. Her stool tested positive for *Clostridium dif-*

From the Department of Surgery, Mount Sinai School of Medicine, New York, New York. Address reprint requests to Joel J. Bauer, M.D., 25 East 69th Street, New York, NY 10021.

© THE MOUNT SINAI JOURNAL OF MEDICINE Vol. 63 No. 2 March 1996

ficile toxin and the patient started to receive vancomycin and flagyl orally. Her symptoms abated after several days and repeat *C. difficile* studies were negative.

The patient was without complaints until 2 months before admission when she began experiencing shortness of breath and chest pain, associated with nonproductive cough and fever. Chest x-ray revealed scattered patches of consolidation with prominent interstitial markings. Bronchial washings were negative for *Pneumocystis carinii* and malignant cells. A transbronchial biopsy exhibited non-necrotizing granulomas. The biopsy was negative for acid-fast bacilli, but positive for intracellular budding yeast cells. A methenanine silver stain identified yeast forms of *Histoplasma*. The patient was treated with amphotericin B and was discharged from the hospital with partial improvement in her respiratory symptoms.

Three weeks before her presentation to the Surgical Service, the patient began complaining of mild abdominal discomfort with a return of blood-tinged diarrhea. Stool was negative for *C. difficile* toxin. Severe cramping epigastric and right upper quadrant pain was exacerbated by oral intake and relieved by vomiting. She noted anorexia, fevers to 102°F, and frequent diarrhea. She became increasingly constipated and, finally, obstipated.

The patient's medical history was significant for two episodes of pelvic inflammatory disease for which she was hospitalized and received antibiotics intravenously. She had had no previous surgery and reported having multiple sexual partners. The patient denied abusing intravenous drugs and using alcohol. She had not traveled outside of New York State.

Physical examination revealed a cachectic-appearing female in obvious distress with a temperature of 101.5°F, pulse of 96, respirations of 24, and blood pressure of 100/70 mm Hg. Findings included decreased breath sounds at both lung bases with bilateral rales; a moderately distended, firm abdomen with no bowel sounds; diffuse tenderness, especially in the right upper quadrant, without generalized rebound or involuntary guarding. The liver edge was palpated 6 cm below the right costal margin with a span of 12 cm. The spleen was not palpable. No other masses were appreciated and the abdomen was highly tympanitic. The patient complained of pain on rectal examination without any evidence of rectal pathology and there was no stool in the rectal vault.

Hematologic values were hemoglobin 8.4 g/dL, hematocrit 24.2%, platelets 187,000/μL. The white blood cell count was 4200 cells/mm^3 with 70% neutrophils, 20% lymphocytes, 5% monocytes, 4% eosinophils. T cell count (CD-4 subset) was 50. Electrolytes were within normal limits, with a BUN/Cr ratio of 12:1. Liver function tests revealed elevated SGOT/SGPT of 56/35 units/L, alkaline phosphatase 170 units/L, total bili/direct bili of 1.5/0.8 mg/dL. The amylase was 210 units/L and the sedimentation rate was 39 mm/hr.

Flat and upright abdominal films showed multiple loops of distended small bowel with multiple air-fluid levels. There was no air under the diaphragm and air and stool were absent from the large bowel. Abdominal CT scan revealed mesenteric adenopathy, hepatosplenomegaly, multiple dilated loops of small bowel with collapsed transverse and descending colon. An upper gastrointestinal series with a small bowel follow-through disclosed a high-grade small bowel obstruction with a hold-up of barium 6 cm distal to the ligament of Treitz.

The patient was prepared for surgery and a laparotomy was performed. Findings included minimal ascites, hepatosplenomegaly, and fatty infiltration of the liver. The small bowel was moderately distended, with three fibrotic, obstructing lesions in the mid-jejunum, the proximal ileum, and the mid-ileum. Three separate small bowel resections were performed. An incidental appendectomy along with a needle biopsy from the right lobe of the liver were also performed. Pathologic study of the small bowel specimens revealed transmural inflammation with thrombosis and scarring. Ulcerations were present on the mucosa with extensive fibrosis extending to the serosa. Externally the jejunal specimen consisted of a circumferential stricture with nearly complete occlusion of the bowel lumen. Microscopic examination revealed an intense inflammatory infiltrate with the formation of granulomas. Methenamine silver stain identified yeast forms of *Histoplasma* adjacent to the ulcerations. Findings on liver biopsy included fatty infiltration evolving to early cirrhosis. The vermiform appendix was unremarkable and negative for fungi by special stains. Postoperatively, the patient did well and continued amphotericin B therapy.

Discussion

Clinical infection with *H. capsulatum* is common in the AIDS population, occurring in 2 to 5% of AIDS victims in endemic areas (e.g., midwestern United States, portions of Central and South

TABLE 1
Clinical Manifestations of Histoplasmosis in AIDS

Common	Uncommon	Rare (<5%)
Disseminated (95%)	Meningitis (15%)	Eye
Fever (95%)	Septicemia (10%)	Pericardium
Weight loss (95%)	Gastrointestinal (10%)	Prostate
Pneumonitis (50%)	Cerebritis (5%)	Pancreas
Hepatosplenomegaly (25%)	Skin lesions (5%)	Pleura
Lymphadenitis (20%)		

Adapted from Wheat J. Infected Dis Clin North Am. 1994;8(2): 467–482.

America), with up to 25% in selected cities such as Indianapolis, Kansas City, Memphis, and Nashville (3). In contrast, less than 1% of AIDS patients are affected in nonendemic areas. In endemic areas, the clinical syndromes are likely to be caused by new or recent infection with *H. capsulatum,* whereas elsewhere it appears that reactivation of latent infection is likely.

In AIDS patients with clinical infection, disseminated disease is extremely common (approximately 95%) (3). Localized pulmonary infection, which is the norm in the nonimmunocompromised population, may occur in patients with CD-4 counts greater than 300. In AIDS, clinical gastrointestinal tract involvement is present in only 10% of patients with histoplasmosis (**Table 1**). Infiltration of the GI tract probably results from hematogenous dissemination of organisms from a pulmonary focus. However, isolated lesions of the gastrointestinal tract have also been reported (5, 6).

Histoplasmosis of the gut leads mainly to intestinal ulceration and localized granulomatous lesions. In cases where the colon is involved, the lesion usually begins as an elevated plaque with progression to a localized granulomatous ulcer. Uncommonly, such foci will assume a different pattern of growth, resulting in a mass that can simulate carcinoma as an "apple core" lesion (2, 6), or as a fungating intraluminal tumor (7). Since Darling (1) reported his original cases of histoplasmosis with multiple mucosal ulcers of the small and large bowel, a variety of pathologic and morphologic lesions of the gastrointestinal tract by histoplasmosis have been reported. These include chronic ulcerative colitis (9), appendicitis and mesenteric adenitis (9), gastric ulcer with fatal hemorrhage (11), inflammatory pseudopolyposis of the small and large bowel (11), recurrent intussusception (12), giant intestinal villi of the small bowel with protein-losing enteropathy (13), esophageal histoplasmosis with mediastinal abscess (13), small bowel obstruction (16), hypertrophic gastric folds (17), and terminal ileitis mimicking Crohn's disease (18).

The symptoms of disseminated histoplasmosis with gastrointestinal involvement are nonspecific. Cappel et al. (18) have documented that in disseminated histoplasmosis, fever, sweats, malaise, and pulmonary symptoms followed by weight loss predominate. With gastrointestinal involvement, however, whether isolated or part of widely disseminated disease, gastrointestinal symptoms predominate but pulmonary symptoms and fever are uncommon. This obviously can delay diagnosis, because the infection is not suspected without fever and pulmonary symptoms. Some of these nonspecific symptoms are poorly defined, usually crampy abdominal pain, diarrhea with or without occult bleeding, nausea and vomiting in 77 documented cases of gastrointestinal histoplasmosis (**Table 2**). Our patient's symptoms included almost all those enumerated, beginning with mild abdominal discomfort with blood-tinged diarrhea progressing to severe epigastric pain associated with nausea and vomiting.

As with other forms of disseminated disease, peripheral lymphadenopathy and hepatomegaly are common physical findings in gastrointestinal histoplasmosis (19). Nevertheless, although clinically obvious gastrointestinal involvement is infrequent, oral lesions are common (20) and can serve as readily accessible biopsy sites. Other reported physical findings in the 77 cases reviewed by Cappel et al. (18) have included abdominal tenderness, rebound tenderness, abdominal distension, and an abdominal mass (**Table 3**). Our patient manifested signs of bowel obstruction, namely abdominal distension with diffuse abdominal tenderness. Of the 29 patients with gastrointestinal histoplasmosis reviewed by Miller and

TABLE 2
Symptoms in 77 Reported Cases of Gastrointestinal Histoplasmosis

Symptom	Cases	
	No.	Percent
Diarrhea	25	32
Weight loss	25	32
Fever	24	31
Abdominal pain	20	26
Lethargy/fatigue	17	22
Dysphagia	15	19
Nausea/vomiting	9	12
Pulmonary symptoms, constipation melena	1–8/each	4–10

Adapted from Cappel et al. Dig Dis Sci. 1986;33:353–360.

TABLE 3

Physical Findings on Admission in 77 Reported Cases of Gastrointestinal-Histoplasmosis

Physical finding	Cases	
	No.	Percent
Peripheral lymphadenopathy	22	32
Hepatomegaly and/or splenectomy	21	30
Abdominal tenderness	12	17
Fever	11	16
Oropharyngeal ulcers or nodules	10	14
Guaiac-positive stool or GI bleed	6	9
Rebound tenderness	5	7
Abdominal distension	3	4
Abdominal mass	2	3

Adapted from Cappel et al. Dig Dis Sci. 1986;33:353–360.

Everett (2), ten had localized granulomatous disease, nine had predominantly ulcer formation, four patients had a pseudopolypoid inflammatory disease, three had diffuse granulomatous reactions, two had evidence of external compression of the gut, and one patient had a calcified, histoplasma-containing cyst in the submucosa of the distal portion of the esophagus. Cappel et al. (18) also reported gross pathologic findings in their review of 77 cases, with ulcerations being the most common finding **(Table 4).** The unusual findings in our patient were three fibrotic, constricting lesions of the small bowel with transmural inflammation and stricture formation. The large bowel was not involved.

Diverse radiographic findings have also been described in patients with gastrointestinal histoplasmosis, but none of them are specific for the disease. Depending on the underlying pathologic changes, six roentgenographic patterns were identified by Lee and Lin (21): 1) malabsorptive (edema, diffuse inflammatory infiltrates, and shallow ulcer), 2) ulcerative (larger and deeper ulcers); 3) polypoid (nodular hyperplasia of the lymphoid follicles and dilated villi); 4) granulomatous (intense and diffuse granulomatous infiltrates); 5) tumor-like (large granulomas); and 6) compressive (enlarged mediastinal or mesenteric nodes). Perforation, hemorrhage, and obstruction are the major complications. Most ulcers are too small and shallow to be demonstrated by barium study, but the associated inflammatory mucosal and submucosal changes are identified as irregular or thickened mucosa in the esophagus, stomach, or bowel. Large ulcers that are deep and circumferential with sharp, elevated edges are characteristic of histoplasmaic ulcers and may be helpful in the differential diagnosis of other ulcerative diseases of the gastrointestinal tract (8). Because the clinical and radiologic findings are so variable, the differential diagnosis of gastrointestinal histoplasmosis is large and may include Crohn's disease, ulcerative colitis, extrinsic narrowing from adjacent neoplastic or inflammatory process, such as amebiasis, tuberculosis, lymphoma, and amyloidosis. Gastrointestinal polyposis, appendicitis, and carcinoma are also included in the differential.

TABLE 4

Gross Pathology in 77 Reported Cases of Gastrointestinal Histoplasmosis

Lesion	No. of cases
Ulcerations	23
Constriction or mass	14
External compression or adhesions	12
Thickened or indurated bowel wall	7
Ulcers with perforation	4
Polyps	3
Fluid-filled cyst	1

Adapted from Cappel et al. Dig Dis Sci. 1986;33:353–360.

The definitive diagnosis of gastrointestinal histoplasmosis depends on culture of the organism in involved tissues, but presumptive diagnosis may be made by evidence of granulomas and yeast in biopsied tissue as demonstrated through methenamine silver stain. If feasible, endoscopy with biopsy and fungal cultures is preferred over exploratory laparotomy for diagnosis.

The mortality of untreated disseminated histoplasmosis is nearly 84% (22). Treatment of gastrointestinal histoplasmosis usually consists of surgical resection and amphotericin B therapy. However, medical management is recommended with surgical intervention reserved for acute emergencies, such as bowel perforation, obstruction, and hemorrhage, or when biopsy is mandatory for diagnosis. Amphotericin B should be administered to immunocompromised patients with obvious signs and symptoms of dissemination with ketoconazole being an effective alternative in immunocompetent patients (23). Hawkins et al. (24) documented a favorable response to ketoconazole with added benefit of no side effects of amphotericin B. In a multicenter, randomized clinical trial involving 54 patients, most with serious underlying disease, 63% were cured with dosages of 400 mg or 800 mg daily (23).

Conclusion

Along with other reported cases, this case reinforces the importance of considering *H. capsulatum* in the differential diagnosis of any AIDS patient who develops gastrointestinal signs and symptoms. The infection in our patient, a New York State resident, manifested itself in an area in which the fungus is nonendemic, which demonstrates the necessity of obtaining an accurate travel history as well as the importance of considering dissemination or reactivation from a pulmonary focus in an immunologically suppressed patient. Initial signs and symptoms are nonspecific, as are the radiographic findings. Diagnosis can be facilitated by a high index of suspicion and an awareness of the ability of the disease to masquerade as other disorders. Diagnosis can be suggested by demonstration of the characteristic yeast cells in a biopsy specimen. The efficacy of therapy with amphotericin B is well established, and this agent should be administered to patients with obvious dissemination or local disease. Surgical intervention should be reserved for patients whose medical treatment has failed or whose condition strongly suggests intestinal obstruction or perforation.

References

1. Darling ST. Histoplasmosis: A fatal infectious disease resembling kalaazar found among natives of tropical Americ. Arch Intern Med 1908; 2:107–123.
2. Miller DP, Everett ED. Gastrointestinal histoplasmosis. J Clin Gastroenterol 1979; 1:233.
3. Wheat J. Histoplasmosis and coccidiomycosis in individuals with AIDS. Infect Dis Clin North Am 1994; 8(2):467–482.
4. Sturim HS, Kouchoukos NT, Ahlvin RC. Gastrointestinal manifestations of disseminated histoplasmosis. Am J Surg 1965; 110:435–440.
5. Weiss ED, Haskell BF. Anorectal manifestations of histoplasmosis. Am J Surg 1952; 84:541–544.
6. Khalil M, et al. African histoplasmosis masquerading as carcinoma of the colon. Dis Colon Rectum 1989; 32: 518–520.
7. Haggerty CM, Britton MC, Dorman JJM, et al. Gastrointestinal histoplasmosis in suspected acquired immunodeficiency syndrome. West J Med 1985; 143:244–246.
8. Henderson RG, Pinkerton H, Moore LT. Histoplasma capsulatum as a cause of chronic ulcerative colitis. JAMA 1942; 118:885–889.
9. Raftert A, Trafas P, McClure RD. Histoplasmosis: A common cause of appendicitis and mesenteric adenitis. Ann Surg 1950; 132:720–728.
10. Fitzpatrick TJ, Neiman BH. Histoplasma capsulatum infection associated with gastric ulcer and fatal hemorrhage. Arch Intern Med 1953; 91:49–55.
11. Bersack SR, Howe JS, Rabson AS. Inflammatory pseudopolyposis of the small and large intestines with the Peutz-Jeghers syndrome in a case of diffuse histoplasmosis. Am J Roentgenol 1958; 80:73–78.
12. Carmona MG, Allen MS. Recurrent intussusception in a six-year old child with histoplasmosis of Peyer's patches. J Fla Med Assoc 1958; 44:955–956.
13. Bank S, Trey C, Gans I, Marks IN, et al. Histoplasmosis of the small bowel with giant intestinal villi and secondary protein-losing enteropathy. Am J Med 1965; 399: 492–501.
14. Jenkins DW, Fisk DE, Walker WE. Esophageal obstruction caused by mediastinal histoplasmosis: Beneficial results of operation. Surgery 1984; 95:59–62.
15. Mohanty SK, Reddy P, Polk JW. Acute small bowel obstruction from disseminated histoplasmosis. Mo Med 1979; 67:100–103.
16. Fisher JR, Sanowki RA. Disseminated histoplasmosis producing hypertrophic gastric folds. Dig Dis Sci 1978; 23:282–285.
17. Alberti-Flor JJ, Granada. Ileocecal histoplasmosis mimicking Crohn's disease in a patient with Job's syndrome. Digestion 1986; 33:176–180.
18. Cappel MS, Mandell W, Grimes MM, et al. Gastrointestinal histoplasmosis. Dig Dis Sci 1986; 33:353–360.
19. Reddy R, Gorelick DF, Brasher CA, et al. Progressive disseminated histoplasmosis as seen in adults. Am J Med 1970; 48:629–636.
20. Goodwin RA, Shapiro JL, Thurman GH, et al. Disseminated histoplasmosis: Clinical and pathological correlations. Medicine 1980; 59:1–33.
21. Lee KR, Lin F. Gastrointestinal histoplasmosis, roentgenographic, clinical and pathological correlation. Am J Gastroenterol 1975; 63:255–265.
22. Fucolow ML, Doto IL, Tosh FE, et al. Course and prognosis of untreated histoplasmosis. JAMA 1961; 177:292–296.
23. Dismukes WE, Cloud G, Bowles C, et al. Treatment of bastomycosis and histoplasmosis with ketoconazole. Ann Intern Med 1985; 103:861–872.
24. Hawkins SS, Gregory DW, Alford RH. Progressive disseminated histoplasmosis: Favorable response to ketoconazole. Ann Intern Med 1981; 95:446–449.

Submitted for publication December 2, 1993.
Final revision accepted February 2, 1996.

Letters

Prostate Cancer Screening

To the Editor: As an Attending Physician at Internal Medicine Associates (IMA) at Mount Sinai, I read with interest the latest issue of *The Mount Sinai Journal of Medicine* (1995;62(4):316–321) devoted to cancer screening. I was particularly pleased to see a study by two of our fourth-year medical students which used IMA as one of the study sites (1). Unfortunately, the authors based their study on a faulty premise, causing them to reach errant conclusions with what is otherwise intriguing data.

The authors stated, ". . . both the American Cancer Society (ACS) and the National Cancer Institute (NCI) have recommended an annual digital rectal examination (DRE) and a prostate-specific antigen blood test (PSA) for screening of prostate cancer beginning at 40 years of age." They went on to state that because of these recommendations, their study sample included all patients older than 40. However, these are, quite simply, not the correct recommendations of the ACS and the NCI. The ACS recommends a DRE for every male patient older than 40 years of age, and a PSA only for men older than 50. The NCI is considerably less convinced of the value of these tests, stating "There is insufficient evidence to establish that a decrease in mortality from prostate cancer occurs with screening by digital rectal examination, transrectal ultrasound, or serum markers including prostate specific antigen." (2) The NCI has recommended DRE, but not PSA for routine screening.

Additionally, the authors failed to mention that a number of other authorities have also decided not to recommend these tests for prostate screening. Neither the United States Preventative Services Task Force nor the Canadian Task Force recommend either DRE or PSA routinely for men. Dr. Michael J. Barry, in a special presentation for the American College of Physicians Annual Session 1995, stated that "such screening should be optional pending results of controlled trials." (3) Finally, in the same issue of the *Journal,* Dr. Alexander Kirschenbaum gave an excellent, concise summary of the current evidence surrounding prostate cancer screening (4). He concluded that not enough data are currently available to definitively decide the usefulness of PSA screening. He suggested two groups of patients who may benefit from screening, after discussion of the risks and benefits. These include men older than 50 years of age with a positive family history of prostate cancer, and men older than 50 with an estimated life expectancy greater than 15 years.

Thus, the jury is still out with respect to prostate cancer screening, although DRE is clearly more widely accepted currently than PSA. IMA is the major ambulatory teaching site for the interns and residents of Mount Sinai Hospital. It is the job of my colleagues and myself to educate the housestaff in the most up-to-date literature, especially regarding such primarily outpatient-oriented issues as cancer screening. The finding that a vast majority of IMA patients surveyed (72%) did receive a DRE, but only a small number (10%) had a PSA done, demonstrates that the correct approach is being practiced, and, in fact, that the housestaff are being properly taught and are following recommended guidelines. It is more interesting that 68% of FPA ("private") patients surveyed had a PSA done. It might be intriguing to study why such a high percentage of these patients had the PSA test, despite the lack of a consensus as to its usefulness.

Todd Simon, MD
Internal Medicine Associates
Box 1087
Mount Sinai Medical Center, New York

1. Perez N, Tsou H. Prostate cancer screening practices: Differences between clinic and private patients. Mt Sinai J Med. 1995; 62(4):316–321.
2. PDQ Supportive Care/Screening/Prevention Information, Screening for prostate cancer, *CancerNet* database, National Cancer Institute, updated July 1995.
3. Barry M. Early diagnosis of prostate cancer. In: Evidence for and against screening: Breast, colon, and prostate, pp. 1–7. Unpublished handout, American College of Physicians Annual Session 1995.
4. Kirschenbaum A. Screening in the new environment of medicine: Prostate cancer. Mt Sinai J Med. 1995;62(4): 257–261.

Dr. Perez Replies: The authors would like to extend their appreciation to Dr. Simon for his comments on and clarification of the prostate cancer screening recommendation by the ACS and the NCI. We stand corrected. As Dr. Simon has pointed out, both groups recommended digital rectal examinations (DRE) for males older than 40 years of age. Because of this recommendation the cut-off age in this study was 40. A retrospective statistical analysis that included only men over the age of 50 produced identical results. There was no difference in the frequency of DRE performed, but the number of PSA tests ordered was statistically significantly different ($p < .01$).

The authors are aware of the current controversy regarding whether population screening for prostate cancer is currently indicated. At present, controlled randomized trials addressing the health benefits of prostate cancer screening are under way in the United States and Europe, but the results will be unavailable for more than a decade (1). Dr. Simon failed to see the main purpose of the study. The purpose of the study was not to determine whether the recommended prostate cancer screening was being practiced at the different study sites but rather to determine the frequency at which prostate screening procedures was being carried out at the different study sites. The authors would like to commend the staff at the IMA clinic for their collaboration in this study. It was gratifying to find that the IMA's predominately low-income population was being screened adequately for prostate cancer.

In conclusion, we also had wanted to address the question of why the PSA test had been performed in a greater number of private patients. One reason seems to be that these patients personally requested that their physicians test their blood for "prostate cancer," a procedure about which they had learned from extensive media coverage. However, it is important that the primary care physician evaluate the medical indications, risks, benefits, and costs of any tests and treatments before providing advice to their patients.

Norvin Perez, MD
New York, New York

1. Gohagan JK, Prorok PC, Kramer BS, Cornett JE. Prostate cancer screening in the prostate, lung, colorectal and ovarian cancer screening trial of the National Cancer Institute. J Urol. 1994;152:1905–1909.

Required Form for Authors:

AUTHORSHIP RESPONSIBILITY, FINANCIAL DISCLOSURE, AND ASSIGNMENT OF COPYRIGHT

Article Title (type):

Each author must read and sign this form dealing with (1) authorship responsibility, (2) financial disclosure and (3) copyright transfer. An author who was a US federal employee when this work was conducted and prepared for publication, must sign statement (4) instead of statement (3).

1. Authorship Responsibility

I certify that I have participated sufficiently in the conception and design of this work, and the analysis of the data (where applicable), as well as the writing of the manuscript, to take public responsibility for it. I believe that the manuscript represents valid work. I have reviewed the final version of the manuscript and approve it for submission for publication. Neither this manuscript nor one with substantially similar content under my (our) authorship has been published or is being considered for publication elsewhere, except as described in a separate attachment. Furthermore, I agree to produce the original recorded data on which the manuscript is based for examination by the Editor-in-Chief and/or his assignees should such records be requested.

2. Financial Disclosure

I certify that I have no affiliation with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials being discussed in the manuscript (eg, employment, consultancies, stock ownership, honoraria), except as disclosed in a separate attachment.

Author(s) Signature(s)	Date	Author(s) Signature(s)	Date
1st		4th	
2nd		5th	
3rd		6th	

3. Copyright

In compliance with the Copyright Revision Act of 1976, the signature of each author on this form shall indicate and be evidence of agreement and mutual understanding with *The Mount Sinai Journal of Medicine*, in consideration of the further action of the *Journal* to review and edit the manuscript being submitted, to transfer, assign, and convey all copyright ownership, including any and all rights incidental thereto, exclusively to *The Mount Sinai Journal of Medicine* if this work is published in the *Journal*.

Author(s) Signature(s)	Date	Author(s) Signature(s)	Date
1st		4th	
2nd		5th	
3rd		6th	

4. US Federal Employees

I was an employee of the US federal government when this work was conducted and prepared for publication; therefore, it is not protected by the Copyright Act and copyright ownership cannot be transferred.

Author(s) Signature(s)	Date	Author(s) Signature(s)	Date
1st		4th	
2nd		5th	
3rd		6th	

Return the original signed form to The Editor, The Mount Sinai Journal of Medicine, Mount Sinai School of Medicine, Box 1094, One East 100th Street, New York, N.Y. 10029. Retain one copy of the form for your file. You may make as many photocopies as needed if your manuscript has more than six authors.

Information for Authors

THE MOUNT SINAI JOURNAL OF MEDICINE, established in 1934 as *The Journal of the Mount Sinai Hospital*, is a peer-reviewed general medical journal. It is indexed or abstracted in *Index Medicus, Current Contents, Science Citation Index, Hospital Literature Index, International Nursing Index, Excerpta Medica, Chemical Abstracts, Biological Abstracts*, and major databases. Theoretical and basic science articles for the clinician, reviews, reports of clinical or research experience, articles on any subject affecting the practice of medicine, short notes in any specialty, book reviews, letters, and articles by medical students are welcome. The *Journal* is a participant in the Agreement on Uniform Requirements for Manuscripts Submitted to Biomedical Journals. The *Journal* aims at responding to authors of manuscripts within 8 weeks of submission.

Communications. Address correspondence on manuscripts review to: Editor, *The Mount Sinai Journal of Medicine*, The Mount Sinai School of Medicine, Box 1094, One East 100th Street, New York, NY 10029-6574. Phone: (212) 241-6108; FAX: (212) 722-6386; e-mail: msjournal@msvax.mssm.edu.

Authorship Responsibility, Financial Disclosure, and Assignment of Copyright. Each author is required to sign (1) the statement on authorship responsibility; (2) the statement on financial disclosure; and (3) either the statement on copyright transfer or the statement indicating that the author is or was a US government federal employee when this work was conducted and prepared for publication. These statements are printed from time to time on a page in the *Journal* entitled "Required Form for Authors" and can be photocopied for use. Copies may also be obtained from the *Journal* office. Use additional copies of the form if there are more than six authors, and replace with the appropriate number of the author, e.g., 7th, 8th, etc. The signed original form(s) should be submitted with the manuscript.

Manuscript Preparation. Submit manuscripts in triplicate, including three sets of illustrations (see below). **Manuscripts must be typewritten or computer generated, double spaced throughout (including references, legends, and tables)**, on one side of 8½ × 11 inch white bond paper with generous margins. All pages except the title page should be numbered consecutively. Begin a new page for the title page, abstract, initial page of text, references, each table, and legends. A floppy disk, if available, should accompany the manuscript. The label on the disk should specify the name of the first author, the title of the manuscript, how the disk has been formatted, and the word processing system used (DOS, Windows, or Windows 95). PC format is preferred but a high-density formatted Macdisk is acceptable. The text should be in ASCII, Microsoft Word or Wordperfect. Other than the title page, the abstract page and each page of text should be numbered sequentially, and include a heading (top right) which identifies the last name of the first author.

Drugs must be identified by their generic names. Do not use abbreviations on the title or abstract page. Define all abbreviations on first use in the body of the manuscript.

Any material taken directly from other sources must be accompanied by a written statement from both the author and publisher giving permission to the *Journal* for reproducing it. The authors must document in the manuscript that all clinical investigations described in the manuscript were conducted in accord with the guidelines of the Declaration of Helsinki and have been approved by the Institutional Review Board or its equivalent. The text must indicate that the informed consent of each subject was obtained. If results of experiments on animals are reported in the manuscript, the text must indicate that these studies were conducted in accord with the National Institutes of Health (US) guidelines for the care and use of laboratory animals and have been approved by the institutional office or its equivalent which supervises such work.

Title Page. The title page should be double spaced, and not numbered. It should indicate (1) title of the article, phrased as concisely as possible; (2) last and first name of each author, middle initials, if any, and not more than two academic degree(s); (3) institutional affiliation and title or rank of each author; (4) name and address of department(s) and institution(s) from which the work originated; (5) the name and address (and FAX number if available) of the author who will receive and respond to correspondence; (6) the name and full postal address of the author from whom reprint(s) are to be requested; (7) acknowledgments of sources of grant support; (8) name, place, and date of any conferences at which some or all of the information was presented; and, (9) a running head of fewer than 40 letter spaces. At the bottom of the title page, *list 3 to 10 key words or phrases* to facilitate indexing in accord with Medical Subject Headings from Index Medicus.

Abstract. All articles must include an abstract of not more than 250 words on the first numbered page of the manuscript. Abstracts should be double spaced and divided into four paragraphs:

Background: what problem(s) was(were) addressed;
Methods: how the study was performed;
Results: the salient results;
Conclusions.

Key Words. Three to 10 key words or phrases must be provided at the bottom of the title page to facilitate indexing. (see above)

Units of Measurement. SI units are preferred, but values may be expressed in conventional units and conversion factors indicated. The range of normal values for the reporting laboratory must be indicated. When conventional units are used in figures or tables, the appropriate conversion factor(s) must be provided in legends or footnotes.

Editing. All papers will be edited to enhance conciseness and clarity without altering meaning and to ensure conformity with journal style. Authors whose primary language is not English should obtain help from someone proficient in English spelling and grammar.

Illustrations. Photographs, line drawings, and graphs should enhance understanding of the text. Three copies of each must be submitted with the manuscript. Line drawings and graphs should be professionally prepared or computer generated and camera ready. Symbols, lettering, and numbers should be large enough to remain easily legible after the figure has been reduced to fit the width of one column. Photographs must be sharp, high contrast, black and white glossy prints not exceeding 5 inches in height × 7 inches in width. The authors must pay attention to the quality of print reproductions of x-ray plates or transparencies to assure that the special features are labeled and will remain easily visible when reduced to fit the width of one column. The name of the subject and hospital number(s) should not be reproduced in the print. Black and white copies of photomicrographs are preferred and must be cropped to a width of 3⅛ inches. The type of stain and power of magnification must be indicated in the relevant legend. Electron photomicrographs should have internal scale markers. On the back of each print and photograph affix a gum label on which the number of the figure has been written to correspond to its identification in the text, title of manuscript, first author's last name, and the word "top" and arrow to indicate the top edge. **Do not write this information directly on the print and do not mount the print on cardboard.** Illustrations which are to appear together should be mounted accordingly as plates and should be equally sized. Protect the prints by enclosing them in cardboard and separate envelopes. **Do not use paper clips.**

Every effort should be made to avoid using a photograph which includes the face or part of the face of a subject. Facial features should be marked to prevent identification of the subject unless the illustration carries details absolutely essential to the text. **In any case, a partial or full facial photograph of a**

subject will not be published unless written permission has been obtained from the subject and accompanies the photograph.

Legends. Legends should not duplicate the text. Each should be on a separate sheet of paper and double spaced. Number each legend to match the illustration it describes. Illustrations mounted as plates to appear together should be accompanied by a single legend identifying the separate elements by position ("upper left," "middle") or by letters. If such illustrations have been labeled with letters, the same letters are to be used in the legend. The legend must indicate the magnification and type of stain used for photomicrographs.

Tables. Each table should be typed on a separate page, double spaced. Number tables consecutively (Table 1, etc.) to correspond to their appearance in the text. Give each table a brief title that is self-explanatory without reference to the text and type the title on a separate line below its number. Identify all abbreviations and/or symbols used in the table in a footnote.

References. All references cited in the text are to be numbered sequentially. Enclose the citation number in parentheses and place it before any punctuation, e.g., Cytomegalovirus (1), steroids (2,3), and recreational drugs (4–6) have all been implicated (7). Do not use superscripts. The reference list must be limited to works cited in the text, and numbered in the order in which they are cited. Type, double spaced on separate pages following the arrangement and punctuation style in the examples below. Journal title abbreviations must conform to those used in *Index Medicus*.

1. Brown LK. Sleep apnea syndromes: overview and diagnostic approach. Mt Sinai J Med 1994; 61:99–112.
2. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. N Engl J Med 1991; 324:424–428.
3. Couser WG. Glomerular diseases. In: Stein J. ed. Internal Medicine 2nd ed. Boston: Little Brown: 1987: 834–861.

List the name of all authors when there are four or fewer. When there are more than five, list the names of the first three, then *et al.*

References to personal communications and unpublished data should not be included in the list of numbered references. Instead, such information should be placed in the text in parentheses. Permission must be obtained from the appropriate individual to so quote the material. This permission must accompany the manuscript at the time of its submission. When the manuscript is accepted for publication, the author(s) will be asked to confirm the accuracy of all references.

Proofs. Proofs are sent to the corresponding author from the printer. Stylistic changes should not be made. Because of increased printing charges, the cost of author's alterations beyond routine correction of typographical errors in text and data will be billed to the author. Proofs should be corrected and returned to and received at the editorial office within 48 hours by express delivery at the author's expense to the address listed above ("Communications").

Reprint Orders. A reprint order form is sent to the corresponding author with proofs; if reprints are desired, the form must be returned to the editorial office, preferably with the proofs, but not later than 3 days to avoid late charges. A check for the exact amount in US dollars should be made out to the order of The Mount Sinai Journal of Medicine. Please indicate on the check the last name of the first author, and the title of the article. If for one reason or another, the reprint order form is not received with the proofs, it may be obtained directly from the editorial office via telephone or FAX. Reprints are available in minimum quantities of 100, at a cost depending on the number of printed pages and whether covers are desired. This information is directly available from the editorial office.

LEVY LIBRARY, MT. SINAI MEDICAL CENTER
3 4805 0410169 3

THE
MOUNT SINAI JOURNAL OF MEDICINE

Forthcoming:

Grand Rounds

Targeted Gene Disruption with Antisense DNA: Research and Therapeutic Applications
Alan M. Gewirtz
University of Pennsylvania School of Medicine

The Supraventricular Tachycardias
M. Noelle S. Langan
The Mount Sinai Medical Center

Peptic Ulcer Disease: Paradigms Lost
Mark Babyatsky
The Mount Sinai Medical Center

Androgen Receptor Mutations in Prostate Cancer
Evelyn R. Barrack
The Johns Hopkins University

Theme Issues

Computers in the Doctor's Scientific Office
Editors: Ray Cornbill and Joseph Kannry

Laparoscopy
Editors: Barry Salky and Michael Edye

Cross-Cultural Issues in Psychopharmacology
Editors: John Herrera and William Lawson

Emergency Medicine
Editors: Andrew Jagoda and Lynne Richardson

Available now:

Maternal and Infant Health: New Perspectives
Kenneth V. Lieberman and
Charles J. Lockwood, editors
volume 61, no. 5, October 1994
88 pages **$15**

Update on Sleep Disorders
Lee K. Brown, editor
volume 61, no. 2, March 1994
96 pages **$15**

Environmental and Occupational Medicine in Clinical Practice
Richard P. Wedeen, editor
volume 62, no. 5, October 1995
96 pages **$15**

The Mount Sinai Hospital 1852 New York

Mount Sinai School of Medicine of the City University of New York

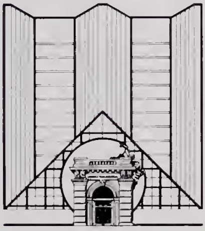
THE MOUNT SINAI JOURNAL OF MEDICINE

VOLUME 63 NUMBERS 3 & 4 MAY / SEPTEMBER 1996

JUN 25 1996

CENTENARY

THE MOUNT SINAI
ALUMNI
1896



Alumni Theme Issue

PART A

STATEN ISLAND UNIVERSITY HOSPITAL
Affiliated with The Mount Sinai Hospital and Mount Sinai School of Medicine

Veterans Affairs

THE MOUNT SINAI JOURNAL OF MEDICINE

The Mount Sinai Journal of Medicine is published by The Mount Sinai Medical Center of New York and has the following affiliates: Bronx Veterans Affairs Medical Center, New York; Elmhurst Hospital Center, Queens, NY; Queens Hospital Center, Queens, NY; and Staten Island University Hospital, Staten Island, NY.

Editor-in-Chief

Sherman Kupfer, M.D.

Editor Emeritus

Lester R. Tuchman, M.D.

Associate Editors

Harriet S. Gilbert, M.D.
Julius Wolf, M.D.

Administrative Assistant

Joseline Ilagan

Assistant Editor for Housestaff

Matthew J. Kapklein, M.P.H., M.D.

Assistant Editors

David H. Bechhofer, Ph.D.
Constanin A. Bona, M.D., Ph.D.
Edward J. Bottone, Ph.D.
Lewis Burrows, M.D.
Christopher P. Cardozo, M.D.
Mark R. Chassin, M.D., MPH, MPP
Joseph S. Eisenman, Ph.D.
Richard A. Frieden, M.D.
Frederick Friedman, M.D.
Steven Fruchtman, M.D.
Paul L. Gilbert, M.D.
James H. Godbold, Ph.D.
Joseph J. Guarneri, Ph.D.
Richard S. Haber, M.D.
Noam Harpaz, M.D.
Tomas Heimann, M.D.
Sheldon Jacobson, M.D., FACEP
Barry W. Jaffin, M.D.
Andrew S. Kaplan, D.D.S.
Samuel Kenan, M.D.
Suzanne Carter Kramer, M.Sc.
Mark G. Lebwohl, M.D.
Zvi Lefkowitz, M.D.
Kenneth V. Lieberman, M.D.
Lynda R. Mandell, M.D., Ph.D.
Steven Markowitz, M.D.
Bernard Mehl, D.P.S.
Leonard E. Meiselas, M.D.
Myron Miller, M.D.
Edward Raab, M.D.
Elliot J. Rayfield, M.D.
Allan Reed, M.D.
Clive Rosendorff, M.D., Ph.D.
Fred Rosner, M.D.
Allan E. Rubenstein, M.D.
David B. Sachar, M.D.
Henry Sacks, M.D.
Ira Sanders, M.D.
Martin H. Savitz, M.D.
Clyde B. Schechter, M.D.
Stuart C. Sealfon, M.D.
Michael Serby, M.D.
Phyllis Shaw, Ph.D.
Alex Stagnaro-Green, M.D.
Barry D. Stimmel, M.D.
Nelson Stone, M.D.
Max Sung, M.D.
Rein Tideiksaar, Ph.D.
Richard P. Wedeen, M.D.

Editorial Board

Barry S. Coller, M.D.
Barry Freedman, M.B.A.
Richard Gorlin, M.D.
Nathan Kase, M.D.
Panayotis G. Katsoyannis, Ph.D.
Charles K. McSherry, M.D.
Jack G. Rabinowitz, M.D.
John W. Rowe, M.D.
Alan L. Schiller, M.D.
Alan L. Silver, M.D.
Alvin S. Teirstein, M.D.
Rosalyn S. Yalow, Ph.D.

The Mount Sinai Journal of Medicine (ISSN No. 0027-2507; USPS 284-860) is published 6 times a year in January, March, May, September, October, and November in one indexed volume by The Mount Sinai Medical Center, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Subscription price: individuals (U.S. and Canada), $75 per year; libraries and institutions (U.S. and Canada), $115 per year; individuals (outside U.S. and Canada), $100 per year; libraries and institutions (outside U.S. and Canada), $140 per year. Single copies $20. New subscriptions begin with the first issue of the calendar year. Send notice of change of address 60 days before its effective date. Second-class postage paid at New York, NY and at additional mailing offices. POSTMASTER: Send address changes to The Mount Sinai Journal of Medicine, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Telephone number (212) 241-6108; Fax (212) 722-6386. Copyright 1996 The Mount Sinai Hospital.

DISCLAIMER: The statement and opinions expressed in the *Journal* by authors and contributors are solely their own and not those of The Mount Sinai Medical Center or The Mount Sinai Journal of Medicine. All advertisements are expected to conform to ethical standards, and their appearance does not imply endorsement or approval by The Mount Sinai Medical Center and/or The Mount Sinai Journal of Medicine of the products with respect to effectiveness, quality, or safety. The Mount Sinai Medical Center and The Mount Sinai Journal of Medicine do not hold themselves responsible for statements made by the authors or contributors, and for any injury to persons or property which may have resulted from ideas, products, and services referred to in articles or advertisements.

COPYRIGHT: Material printed in the Journal is protected by copyright. In accord with U.S. Copyright law, no part of the Journal may be reproduced, displayed, or transmitted in any form, electronic or mechanical, including photocopying or by any information storage or retrieval system, without prior written permission to do so. Permission should be requested from The Editor. The Mount Sinai Journal of Medicine, Mount Sinai School of Medicine, Box 1094, One East 100th Street, New York, NY 10029.

THE
MOUNT SINAI JOURNAL
OF MEDICINE

Volume 63
Number 3 & 4
May/September 1996

CONTENTS

Commentary *Albert M. Lefkovits* 143

The Nature of Rheumatogenic Streptococci *Gene H. Stollerman* 144

The Emergence of "Emerging Diseases": A Lesson in Holistic Epidemiology *Edwin D. Kilbourne* 159

Research in Psychodynamic Psychotherapy *Robert S. Wallerstein* 167

Renal Biopsy: Why and When *Lewis Reisman, Steven Dikman, Jacob Churg, and Sherman Kupfer* 178

A History of Immunosuppressive Drugs in the Treatment of Inflammatory Bowel Disease: Origins of the Mount Sinai Hospital *Burton I. Korelitz and Daniel H. Present* 191

Pathogenesis of Inflammatory Bowel Disease *Lloyd Mayer, Xian Yang Yio, Annica Lin, and Asit Panja* 202

Calcification of Elastic Fibers in Pseudoxanthoma Elasticum *Sharada Truter, Jodi Rosenbaum-Fiedler, Allen Sapadin, and Mark Lebwohl* 210

Meta-analysis: An Update *Henry S. Sacks, Dinah Reitman, Daniel Pagano, and Bruce Kupelnick* 216

Hemophilia: Yesterday, Today, and Tomorrow *Louis M. Aledort* 225

Prevention into the 21st Century *Arthur L. Frank* 236

Using Standardized Patients for Assessing Clinical Performance: An Overview *Mark H. Swartz and Jerry A. Colliver* 241

Breast Cancer Chemotherapy after Fifty Years: More Cures and Less Mutilation *Ezra M. Greenspan* 250

Hypertension: A Personal Odyssey *Milton Mendlowitz* 256

History of the Department of Dermatology of The Mount Sinai Hospital *Douglas D. Altchek* 263

Vena Cava Filters: Available Devices and Current Practices *James M. Cooper, James Silberzweig, and Harold A. Mitty* 273

THE MOUNT SINAI
ALUMNI
1896

Commentary

Albert M. Lefkovits, M.D., President

The Mount Sinai Alumni

This issue of The Mount Sinai Journal of Medicine commemorates the Centennial of The Mount Sinai Alumni. It features papers written by some of Mount Sinai's most distinguished graduates and covers a wide range of medical disciplines. The quality of these presentations reflects on the standards for admission to both the undergraduate and graduate medical education and training programs offered by Mount Sinai. It also is a testament of the esteem given to those who are willing to share their knowledge, experiences, and insights gained at the frontiers in the never-ending effort to promote the best in research and education for which our institution can be justifiably proud.

Although there are many uncertainties confronting us as physicians in the present climate, we are privileged to practice a profession which is still widely recognized for its nobility, admired for its intellectual accomplishments, and most important, affords all of us the opportunity to experience personal satisfaction and fulfillment.

As the *Alumni* enters into its second century, we can be confident that Mount Sinai will continue to provide the very best in medical education and training for physicians, medical scientists, and allied health workers. A fertile environment is essential to expand the frontiers in both the basic and clinical sciences if delivery of high-quality patient care is forever to be assured.

The Nature of Rheumatogenic Streptococci

Gene H. Stollerman, M.D.

Rarely has a clinical investigator been accorded the privilege to study a major disease continuously for more than half a century. I owe my opportunity to do so primarily to my mentors, George Baehr, Gregory Shwartzman, and Paul Klemperer, who fired my enthusiasm for investigating the inflammatory diseases of connective tissues and their relation to infection during my clinical training at Mount Sinai Hospital. In 1949, upon completion of my chief residency, these eminent medical scientists sent me to the renowned microbiology department of Colin MacLeod at New York University. Dr. MacLeod promptly set me to work in the laboratory of Alan Bernheimer who was studying and purifying streptococcal toxins. It was my assignment to find out whether these had anything to do with the pathogenesis of rheumatic fever (RF). The outcome of that assignment provides the background for this discussion which will trace the development of the current concept of the variation in the structure of the strains of group A streptococci (GrAS) which account for their differences in virulence and for the changing epidemiology of the clinical syndromes they produce. Special emphasis will be placed on RF because modern molecular genetics has brought us to the threshold of the genetic engineering of a vaccine that might prove sufficiently safe and effective to eradicate RF and rheumatic heart disease from the world.

Establishing the Group A Streptococcus as the Sole Cause Of Rheumatic Fever

By the mid-20th century, little doubt remained that only one bacterial species, the group A streptococcus (*Streptococcus pyogenes*), caused rheumatic fever (1). By the 1930s Rebecca Lancefield had characterized GrAS serologically by its cell wall carbohydrate antigen. She further identified the heterogeneous surface M protein as the major virulence factor responsible for resistance to phagocytosis and as the antigen responsible for serologic type specificity. She thereby explained the type-specific basis for human immunity and human susceptibility to multiple episodes of streptococcal sore throat during a lifetime (2).

In the 1930s, Alvin Coburn's clinical epidemiologic studies (3), using Todd's antistreptolysin O titer (4) as a marker of antecedent streptococcal sore throat, made it clear that GrAS infection was the major cause of RF. In that decade, the introduction of the sulfonamides prompted their experimental use in the prevention of rheumatic recurrences by Coburn and Moore (5) and Thomas and France (6). The great World War II RF epidemics among military recruits and the discovery of penicillin in the 1940s set the stage for the elegantly controlled studies at the Warren Air Force Base in Wyoming by Charles Rammelkamp and his associates (7). These left little doubt that initial (primary) attacks of RF could be prevented by prompt and proper penicillin therapy of streptococcal pharyngitis.

The introduction to clinical medicine of the long-acting repository salt of benzyl penicillin, benzathine penicillin G, in the early 1950s (8) (**Fig. 1**) provided the opportunity to develop a safe method of ensuring almost complete and contin-

---

Professor of Medicine and Public Health, Emeritus, Boston University. Address reprint requests to Gene H. Stollerman, M.D., 10 Willow Spring Lane #21, Hanover, NH 03755.

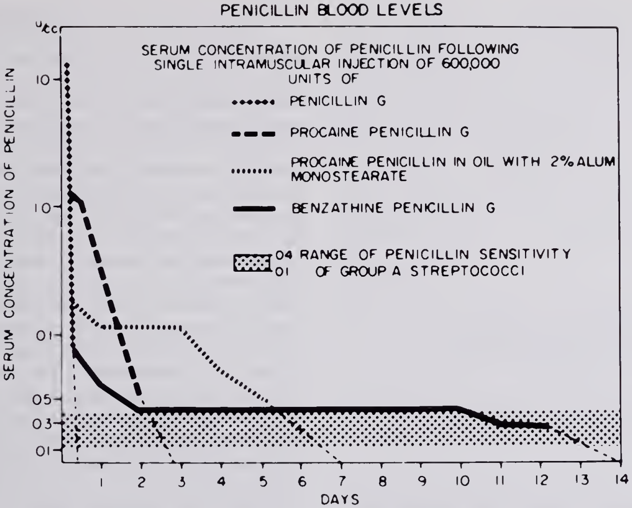


Fig. 1. Mean serum levels of penicillin in patients who received one intramuscular injection of 600,000 U of aqueous penicillin G compared with single injections of three different repository penicillin salts (8).

uous protection of rheumatic subjects against recurrent (secondary) attacks of RF by administering monthly intramuscular injections of the repository compound over several years (9). In randomly controlled prospective studies, such protection proved more effective than prophylaxis with either oral penicillin or sulfonamides for the prevention of rheumatic recurrences (10) and became the preferred method of secondary prophylaxis for high-risk rheumatic subjects **(Table 1).** Moreover, a single injection of benzathine penicillin G for treatment of GrAS sore throat also proved highly effective in preventing primary rheumatic attacks (11), and mass prophylaxis with a single injection of this compound terminated violent epidemics of streptococcal pharyngitis and RF in military populations (12) **(Fig. 2).** Indeed, the liberal use of benzathine penicillin G in the treatment of sore throat was equally effective in civilian populations, as demonstrated by the experience of the tiny Central American country of Costa Rica (13) **(Fig. 3).**

TABLE 1

Summary of Five-Year (1954–1959) Irvington House Study of Three Prophylactic Regimens to Prevent Recurrence of Rheumatic Fever

	Bicillin	Oral penicillin	Sulfadiazine	Total
Patient years (pt. yr.)	560	545	576	1,681
Strep. infections				
(Rate per pt. yr.)				
Total	6.0	20.6	23.2	16.6
Confirmed*	4.3	18.5	17.7	13.5
Reoccurrences				
Number	2	30	16	48
Rate				
% per pt. yr.	0.4	5.5	2.8	2.9
% per strep. infect.	5.9	25.5	11.6	16.5
% per confirmed* strep. infect.	8.3	28.7	15.7	20.7

* Antibody rise.
Source: Ref. 10.

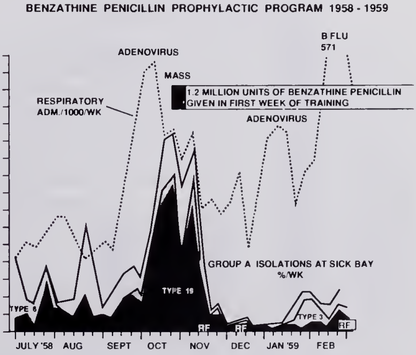


Fig. 2. Mass prophylaxis of RF in a cohort of naval military recruits given 1.2 million U of benzathine penicillin G during an epidemic of type 19 group A streptococcal pharyngitis (12).

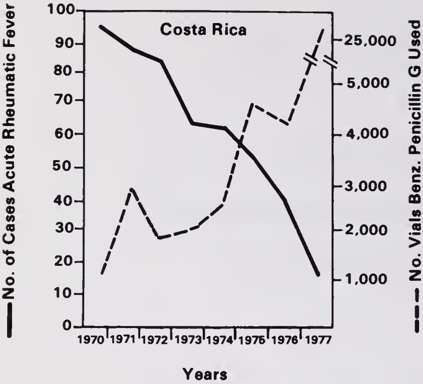


Fig. 3. The increasing use of injectable benzathine (Benz.) penicillin G during the 1970s in Costa Rica and the dramatic decline in the prevalence of RF. (From Ref. 13.)

With almost perfect protection from intercurrent streptococcal infection provided by monthly injections of benzathine penicillin G, and with serial streptococcal antibody determinations to confirm freedom from subclinical (inapparent) GrAS infections, it became possible to determine the cardiac sequelae of first attacks of RF and the subsequent course of rheumatic heart disease in the absence of recurrent rheumatic attacks. Moreover, it was established that all recurrent rheumatic manifestations, including Sydenham's chorea (14, 15), were caused by GrAS infections alone and not to other precipitating factors (e.g., other infections, trauma, psychic insults, etc.).

"Nonrheumatogenic" Group A Streptococcal Infections

Despite the overwhelming evidence, however indirect, that GrAS was the sole etiologic agent of RF, it was obvious to many investigators that not all infections with this species were potentially rheumatogenic. The absence of RF as a sequel of extrapharyngeal infections (pyoderma, wounds, puerperal sepsis, and even those producing the most severe septic states) was well appreciated. In particular, students of post-streptococcal acute glomerulonephritis noted that both sequelae, RF and acute glomerulonephritis (AGN), rarely occurred simultaneously in the same patient after a single episode of streptococcal sore throat (16). What, then were the differences between the infecting strains of streptococci, or was it primarily the host factor that determined sequelae of the infection?

Factors Determining the Attack Rate of RF

One of the most intriguing observations made during the Warren Air Force Base studies in Wyoming was the attack rate of RF after streptococcal pharyngitis in the control group who were not treated with antibiotics. Approximately 3% of these individuals developed RF after streptococcal pharyngitis and these infections were caused by strains belonging to several different M serotypes that were prevalent simultaneously on that base. The attack rate varied from 1 to 5% depending on the antigenic intensity of the infection as measured by antistreptolysin O titers (17). The rates were consistent regardless of age, race, or ethnic group. Host factors clearly related to this response could not be identified (18). Did these high attack rates reflect the virulence of the organism or was there also a qualitative difference in RF-producing and non-RF-producing strains?

Some observers of civilian cohorts of school children, including myself, found it difficult to believe that RF could occur with so high an attack rate considering that virtually all children sustained at least several attacks of GrAS sore throat in a lifetime. That would have made the incidence of rheumatic heart disease in the population enormous. The issue arose as to whether "streptococcal sore throat" was simply overdiagnosed by being mistaken for the much more frequent viral pharyngitides in patients who were carrying, but were not truly infected by GrAS. In addition, were truly rheumatogenic strains not only quantitatively excessively virulent but qualitatively different from nonrheumatogenic strains as well?

In the late 1950s, Alan Siegel et al. (17) studied this issue in Chicago school children. The authors concluded that under nonepidemic conditions many children carrying GrAS could have coexistent viral infection causing many false-positive diagnoses of "streptococcal pharyngitis." Nonetheless, these prospective studies still showed a RF attack rate far below that reported in the Warren AFB studies in the patients whose pharyngitis and GrAS-positive cultures were followed by a significant increase in streptococcal antibodies, unmistakable evidence of invasive streptococcal infection. Also, compared with the Warren AFB epidemics, the relatively benign throat infections in school children were not only clinically milder but the streptococcal strains isolated from them were less virulent and frequently nontypeable. Indeed, the only two cases of RF observed in these prospective studies were caused by highly encapsulated M5 and M3 (mucoid) strains (18, 19).

Nonrheumatogenic Streptococcal Serotypes

As early as 1935, Coburn and Pauli (20) suggested that some strains of GrAS were not rheumatogenic. They were particularly impressed with the rarity of acute rheumatic fever (ARF) in M4 and M28 infections. In 1942 Kuttner and Krumwiede (21) reported the absence of RF after a M4 outbreak of streptococcal sore throat at Irvington House, a hospital for children convalescing from RF. The absence of RF after M12 pharyngitis is particularly noteworthy. Although M12 strains have been associated with epidemic AGN, as have certain other serotypes (22), no documented epidemic of RF has followed M12 infection, nor infections by other common pharyngeal serotypes, such as M2 and M4 (23). As early as 1942, Bloomfield and Rantz (24) observed an explosive, clinically severe food-borne outbreak of streptococcal pharyngitis in more than 300 army recruits without a single confirmed case of RF. That strain belonged to Griffith's serotype 15. A military milk-borne epidemic due to an M12/T10 strain was also noteworthy for the absence of either RF or AGN (25). Such a benign outcome was credited to appropriate diagnosis and treatment. This report prompted my editorial commentary that GrAS strains required specific properties to be either rheumatogenic or nephritogenic. They were rarely if ever both and, indeed, were most frequently neither (26)!

Nephritogenic Streptococci and Streptococcal Pyoderma

After the report by Rammelkamp, Weaver, and Dingle (22) of the 1953 epidemic of AGN at the Bainbridge Naval Base, which was clearly due to a pharyngeal M12 strain, the concept of distinctly nephritogenic strains of GrAS belonging to certain M serotypes seemed to be established. The studies of Parker and his associates in Trinidad (27) and of Top et al. (28) and Wannamaker (29) in Minnesota in the 1960s firmly established epidemiologically the existence of GrAS strains that had primary tropism for the skin and only secondary affinity for the throat. These so-called pyoderma or "skin" strains belonged to certain M and T serotypes that were associated with epidemic AGN (29), but as studies in Trinidad revealed, they were quite different than the strains that were associated with RF on that island. The latter were predominantly M6 and M11 (30). The latter serotype, it should be noted, was positive for the streptococcal "opacity factor," a lipoprotein lipase which is a common property of pyoderma strains but very rarely associated with rheumatogenic strains (31, 32). Widdowson et al. (33) and Maxted and Falkenberg (34) showed that some strains had a great tendency for antigenic drift in some subserotypes, making the relationship of this specific M type to rheumatogenicity unclear. Despite the multiplicity of M and T serotypes in any community at any one time, we proposed that "microepidemics" were due to a rheumatogenic strain giving rise to focal, sporadic cases of RF (19).

In Memphis in the late 1960s, Bisno and colleagues (35) studied the seasonally disparate epidemiology of pyoderma-associated AGN and RF in the southern United States to demonstrate the nonrheumatogenicity of the GrAS strains causing the former sequelae (Fig. 4). Streptococcal pyoderma and pyoderma-associated AGN were shown to be primarily warm weather diseases. RF was virtually absent during summer months despite the frequent occurrence of pharyngitis due to strains of GrAS that belonged to a variety of newly identified M serotypes with properties similar to pyoderma strains (36). When RF reappeared in the late fall, pyoderma, AGN and the associated pharyngeal strains had disappeared. Unfortunately, the patients admitted to the Memphis Medical Center with RF at that time came from a wide referral base. By the time the patients arrived they no longer carried GrAS in their throat. Our attempts to identify the antecedent infecting strains were thus frustrated. Moreover, RF at this time (the 1970s) began to

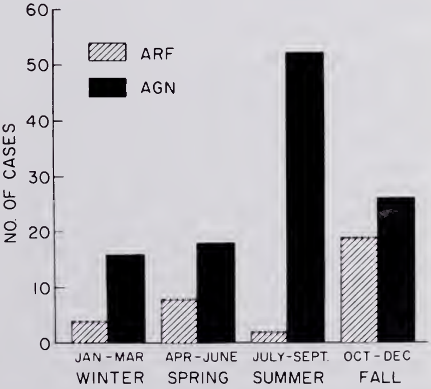


Fig. 4. Dissociation of epidemiology of acute rheumatic fever and acute glomerulonephritis in admissions at City of Memphis Hospitals, September 1965 to August 1968.

disappear spectacularly in our population and elsewhere (37).

The Decline and Reappearance Of RF in Relation to GrAS Strain Variation

Actually, the incidence of RF had been declining in the world for some time. A 100-year record (1860–1960) of the annual incidence of RF in Copenhagen, Denmark suggests that factors other than penicillin may have been operating to modify the disease before the discovery of antibiotics (18). Among such factors were improved living conditions, which reduced the spread of streptococcal sore throat. The most spectacular decline in the incidence of the disease, however, was observed in the 1970s, as documented in the records of the Chicago Health Department's Heart Disease Control Program (38) **(Fig. 5)**. By the 1980s, RF had virtually disappeared in the United States and in similarly developed countries, such as Japan and Australia, at a pace that could not be correlated with socioeconomic or genetic changes alone (38). Despite the disappearance of RF, streptococcal sore throat continued to occur very frequently among school children. Indeed, the frequency of positive streptococcal cultures in school children with pharyngitis occurs at the same high incidence now as it did in the 1950s and 1960s **(Table 2)**.

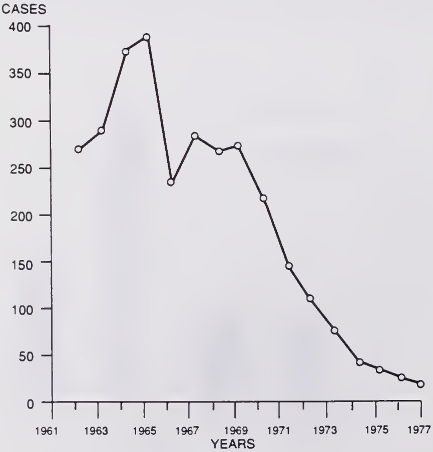


Fig. 5. Decline in number of patients with RF or rheumatic heart disease registered annually in Chicago Board of Health's RF registry between 1965 and 1977 (38).

TABLE 2

Rate of Positivity of Throat Cultures for Group A Streptococci in Rhode Island School Children with Pharyngitis in the Early 1980s

Age, yr	Total no. of cultures	Total no. of positives	% Positive
0–4	3,704	546	14.7
5–9	3,713	1,256	13.8
10–19	5,495	852	15.5
20–39	4,457	425	9.5
≥40	917	55	6.0
?	2,864	464	16.2
Total	21,150	3,598	17.0

Source: Ref. 39.

The Resurgence of RF in the United States and the Reappearance of Rheumatogenic Strains of GrAS

The startling reappearance of RF in the United States in the 1980s took the medical profession by surprise. Outbreaks of the disease appeared without warning in widely separated military populations and in relatively well-to-do civilian communities (40) **(Table 3, Fig. 6)**. Several features of the epidemiology of the outbreaks studied are noteworthy. The cases of RF all had the classical major and minor features of RF, including carditis and chorea (41). In civilian populations the mean age was 11.8 years. These populations were clearly middle class, 94% of the patients were white, 4% were Pacific Islanders, 1% were Hispanic, and 1% were Asian. The mean household size was 6.1 (greater than that of the population base), and mean income was $25,000 to $30,000. (The poverty level for family of six was approximately $16,000.) Fifty-six percent of the patients' parents attended college (42).

When the streptococcal isolates were investigated in the affected military and civilian populations, heavily encapsulated strains ("mucoid" colony formers) of M18 and M3 serotypes were frequently isolated, both notorious serotypes previously associated with epidemic RF and strains

TABLE 3

Epidemic of Rheumatic Fever in Utah, 1985–1986

Major manifestations	No.	%
Carditis	14	14
Polyarthritis	14	14
Chorea	4	4
Carditis and polyarthritis	43	44
Carditis and chorea	14	14
Carditis, chorea, and polyarthritis	6	6
Polyarthritis and chorea	4	4
Total	99	100

Source: Ref. 42.

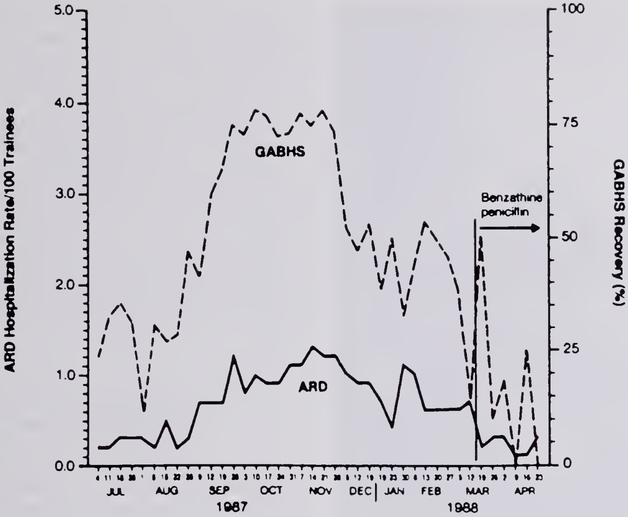


Fig. 6. Hospitalization for acute respiratory disease (ARD) (lower curve, left ordinates) and present recovery of group A streptococci (GABHS) (upper curve, right ordinates) among Army trainees at Fort Leonard Wood, Missouri, July 1987 to April 1988. Thirteen cases of ARF were diagnosed between October 1987 and February 1988. No new cases had been reported since the second week of March when benzathine penicillin G was given once to all nonallergic soldiers and thereafter to new trainees on arrival at the base (Centers for Disease Control. MMWR 1988;37:519).

which had not been seen in several decades (42, 43) (**Fig. 7**). Subsequent surveys confirmed that these mucoid strains that belonged to well-known RF-associated M types had reappeared in areas where focal outbreaks had occurred (44).

The Prevalence of Severe, Invasive GrAS Infections, Streptococcal Toxic Shock Syndrome, and Necrotizing Fasciitis (Streptococcal Gangrene)

Shortly after the resurgence of RF, unusually severe invasive streptococcal infections began to be reported both in the United States and abroad (45, 46) (**Fig. 8**), especially life-threatening infections caused by M1 and M3 serotypes. The point of entry for these infections was usually the skin and soft tissues. Some of these resulted in a toxic shock syndrome and multiorgan failure, with manifestations similar to the staphylococcal toxic shock syndrome. This serious GrAS complication was therefore dubbed "streptococcal toxic shock syndrome" (47). These events stimulated a reinvestigation of the erythrogenic (pyrogenic) toxins that had been purified and studied carefully in the 1950s and 1960s by Kim and Watson (48), but in which interest had become dormant after scarlet fever became a mild and relatively rare disease (49, 50).

The "new" GrAS strains were restudied in the laboratories where Watson's studies of erythrogenic toxins had originated and had been continued by Schlievert and his associates (47). The strains from patients with toxic shock were analyzed for production of (at least) three pyrogenic exotoxins, A, B, and C, all known to be associated with scarlet fever. In contrast to strains producing B and C toxins and only mild scarlet fever, the "new" U.S. strains associated with streptococcal toxic shock produced the A toxin regardless of M type. This toxin was found to share biologic activity and amino acid sequence homology with the *Staphylococcus* pyrogenic toxins associated with the staphylococcal toxic shock syndrome (51). In Sweden, however, strains of M1 elaborating exotoxin B rather than A have predominated (52), and two new pyrogenic exotoxins, "SSA" (53) and

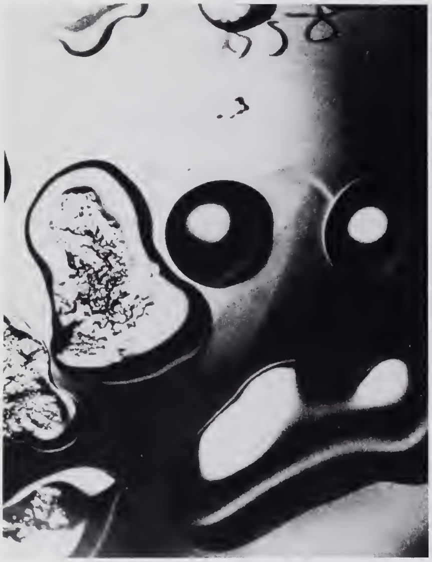


Fig. 7. Low-power micrograph of encapsulated, mucoid GrAS colonies growing on blood agar (61).

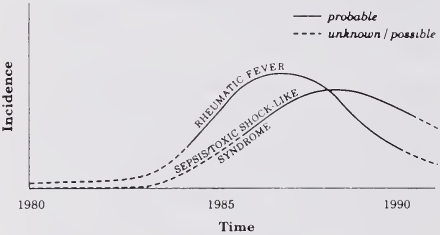


Fig. 8. A schematic representation of the increase in cases of RF and severe GrAS infections (including toxic shock-like syndrome) in the United States during the latter half of the 1980s and early 1990s (46).

"mitogenic factor" (54), were isolated recently from strains of GrAS.

At the beginning of the 20th century, the epidemiology of scarlet fever was closely related to streptococcal pharyngitis and its sequels. The associations between scarlet fever, RF, and AGN were demonstrated clearly by Schick's studies of "The nachkrankheiten of scarlet fever" in 1907 (55). Rheumatogenic strains, however, were not necessarily strong pyrogenic toxin producers *in vivo,* inasmuch as the epidemiologies of the two conditions were not congruent (56). Streptococcal pyrogenic exotoxins and a somatic antigen that is part of the M protein molecule (see below) have both received considerable attention for their "superantigen" effect as activators of T lymphocytes.

In addition to STSS, the incidence of *necrotizing fasciitis* (Meleney's streptococcal gangrene) seems to be increasing as well (57). However independent of each other these streptococcal diseases may be, there seems to be an increasing trend toward the spread of virulent strains of organisms with special potential for host tissue damage throughout North America, Europe, and Australia (57). Why there should be such an intercontinental drift toward virulence in GrAS strains is an intriguing and unanswered question.

The Characteristics of Strains of Rheumatogenic Streptococci

As noted above, the most rheumatogenic strains are very rich in M protein and possess the very large hyaluronate capsules that form "mucoid" colonies on blood agar plates. Such strains are also highly virulent in mice (2) and grow in very short chains in young blood-broth cultures (58). These properties have been observed in *all* strains that I have studied which have caused rheumatic fever epidemics at Warren Air Force Base, at the Great Lakes Naval Training Center, and at institutions, and occasionally antecedent infections of individual patients who developed RF [see Stollerman et al. (19)]. In the early 1970s, Widdowson et al. (33) pointed out that rheumatic fever-producing strains sent to them for serologic identification and typing at the Central Public Health Laboratory in London contained a large amount of acid-extractable M protein compared with other GrAS strains. They called this material "M-associated protein," recognizing its relative impurity. They also pointed out the strong type-specific antigenicity of such strains in humans and noted that they lacked the "opacity factor," lipoprotein lipase usually found in pyoderma strains (31, 32, see above).

Why have so few others pursued these leads? Such strain virulence properties are difficult to study retrospectively. In culture media, these strains dissociate rapidly with loss of capsules and M protein, necessitating passage through mice or through fresh human blood to maintain the virulent phase (2, 58, 59). During convalescent carriage such rapid dissociation usually results in throat carriage of strains of little or no virulence (59, see below). On the other hand, when virulent strains are introduced into a cohort living in close contact, such as families, military recruits, school children, or institutionalized individuals, secondary spread is extremely rapid (60).

It is not surprising, therefore, that these virulence properties are rarely reported in most clinical epidemiology studies. Strains harvested routinely by clinical microbiology laboratories and sent to reference laboratories for serologic testing may lose virulence in the handling. Few clinical laboratories have personnel specially trained to observe colony morphology characteristics, and only a few reference laboratories have appropriate serotyping sera or the resources or interest to restore the virulence properties of strains by frequent mouse passage or passage through fresh human blood (2, 58, 61). Epidemiologic data, moreover, are confused by studies of strains which, although still M-typeable, may not be in a sufficiently high phase of virulence in the community to produce RF. Thus, doubt may be cast on the putative "rheumatogenicity" of a given M serotype. Nor is it clear whether once they become attenuated such strains can readily recover their virulence properties naturally, even during frequent passage through human contacts. Does the appearance and spread of epidemic strains result then, *de novo,* through mutation, bacteriophage infection, or plasmid-mediated gene transformation? Or, in addition, are we importing virulent strains from endemic foci, in the United States and abroad? Military recruit centers have long been the epidemiologic model for the spread of strains introduced by recruits from widespread urban centers of the United States.

In any case, a review of the English language medical literature by Bisno in 1980 (23) revealed that only a limited number of M serotypes have been associated with epidemic RF. These include (not a complete list) M types 5 (most commonly), 3, 6, 14, 18, 19, 24, 27, and 29. Many prevalent M serotypes, including 2, 4, and 12, have been conspicuously absent from this list. It is noteworthy that rheumatic recurrences have all but disappeared except in countries still plagued by RF (62). Recent studies from Santiago de Chile, for example, showed that in local cohorts of school children who were maintained on careful penicillin prophylaxis, rheumatic recurrences disappeared even after prophylaxis was discontinued. Rheumatogenic strains of streptococci also disappeared in this locale, although RF still occurred in other sections of the city or in near-by suburban towns (63). If, indeed, the number of M serotypes causing RF in the world is relatively limited, a safe, multivalent M-protein vaccine would be extremely valuable, particularly to prevent recurrences in rheumatic subjects. What are the prospects?

Molecular Biology of M Protein, Rheumatogenicity, and Prospects for a Vaccine

As early as the 1930s, Rebecca Lancefield recognized the vaccine potential of the M protein that she extracted with hot acid (2). Subsequently, Fox and associates (64), Potter et al. (65), and others attempted to immunize with this partially purified product. Although preparations with adjuvants appeared capable of raising type-specific antibodies, the intense skin hyperreactivity to them in everyone tested was discouraging (66). Moreover, the fluorescence-labeled antibody studies of Kaplan and Svec (67) and Zabriskie and Freimer (68) demonstrated the potential of M protein for cross-reactivity with human tissues, and especially with the myocardium. Kaplan localized this "M-like" myocardial cross-reactive moiety in or very near to the M-protein molecule itself.

By 1973, Beachey and Stollerman (69) had studied further the nature of the cellular toxicity of the hot acid-extracted M protein. Even this relatively "pure" M protein was toxic to all cells exposed to it in human plasma suspensions, as well as to human skin. If separable at all, the type-specific M moiety would have to be split from the toxic moiety on the same molecule, and from tissue cross-reactive moieties as well, if further attempts at producing an M protein vaccine were to be made (18, 70, 71).

Using an earlier suggestion by Rebecca Lancefield, we used gentle extraction of M protein with dilute pepsin at the suboptimal pH of this enzyme, which yielded a remarkably pure type-specific moiety. Itzhak Ofek undertook with Beachey a study of this M-protein fragment's potential as a ligand for the adherence of GrAS to mucosal surfaces. Surprisingly, dilute pepsin removed the type-specific segment of M protein without affecting mucosal adherence (72, 73). Rather, adherence was blocked specifically by a purified lipoteichoic acid. Just as striking a finding, however, was the purity of the M protein piece that had been split by dilute pepsin from the streptococcal surface, leaving behind apparently structurally intact "fimbriae" (74) (**Fig. 9**).

The Molecular Structure of "PepM"

In a relatively brief time, Beachey and his associates (75, 76) produced a series of seminal

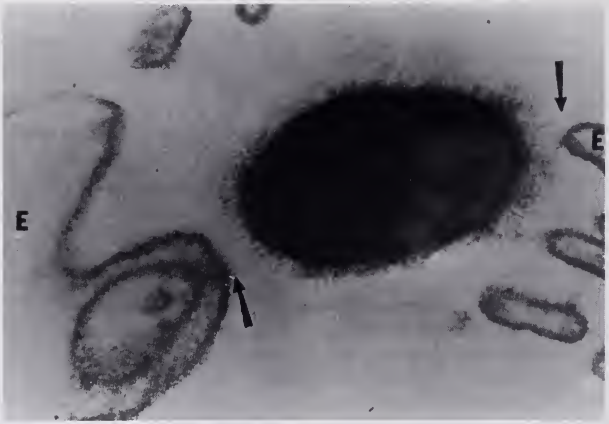


Fig. 9. Electron microphotograph of GrAS adhering to the surface of a buccal mucosal cell. Prominent surface projections (fibrillae) contain both M protein and lipoteichoic acid (72).

studies providing great insight into the molecular structure of the "pep-M24" that had been extracted gently from the surface of M24 streptococci. PepM was revealed to be composed of a series of repeating peptides at the amino-terminal locus that contained the immunogenic type-specific peptide fragment of M24 (**Fig. 10**). It was promptly established as immunogenic in animals and proved to be well tolerated in humans in doses sufficient to evoke bactericidal type-specific antibodies (76). The toxic moieties of M protein apparently having been removed, attention turned to the problem of freeing the M peptides from the potentially "rheumatogenic" epitopes cross-reactive with myocardium and other host tissues, such as tropomyosin, myosin, and vimentin (67, 68, 77–80). Interestingly, these were all elastic, coiled molecular structures, like M protein.

Meanwhile, at the Rockefeller University, Fischetti and his colleagues were engaged in the structural analysis of M proteins and also in their potential for immunization (81–83). They showed the molecule to be an alpha-helical coiled coil structure and confirmed that the amino-terminal regions were hypervariable and could account for type-specific immune responses (79–81). Their studies revealed further that there were two distinct classes of M protein, Class I from strains associated with rheumatic fever epidemics, which had the long repeating sequences of amino-terminal peptides, (**Fig. 11**), and Class II, which were associated with skin infections and the production of "opacity factor" and which had smaller and structurally different M molecules (82, 83).

James Dale addressed the problem of separating the type-specific opsonic epitopes of a highly rheumatogenic type 5 M protein from epitopes cross-reactive with human myocardium that he had found in the pepM-terminal peptides of several heterologous types (M6, M18, and M19). These seemed to be contained within the covalent structure of the M5 fragment itself (78). However, a synthetic peptide made by copying the amino-terminal 20 amino acids of M5 (84) was shown to be immunogenic when coupled to tetanus toxoid; moreover, it was free of cardiac cross-reactivity. This crucial observation proved that the opsonic and tissue cross-reactive epitopes were located within different parts of the same molecule. Additional studies with M6 and M24 synthetic copies produced the same results (85, 86). Moreover, the autoimmune epitopes did not appear to be immunogenic when native rather than synthetic primary structures were used as immunogens.

Localization of Tissue-Cross-reactive M Protein Epitopes

By 1984 Manjula and Fischetti (80) had determined the entire covalent structure of the pepM5 molecule, and by 1986 Dale and Beachey were able to locate the sequence 84–116, which inhibited the myosin-cross-reactive antibodies evoked by pepM5 in rabbits, between the A and B repeating peptides of M5 (87). Other epitopes

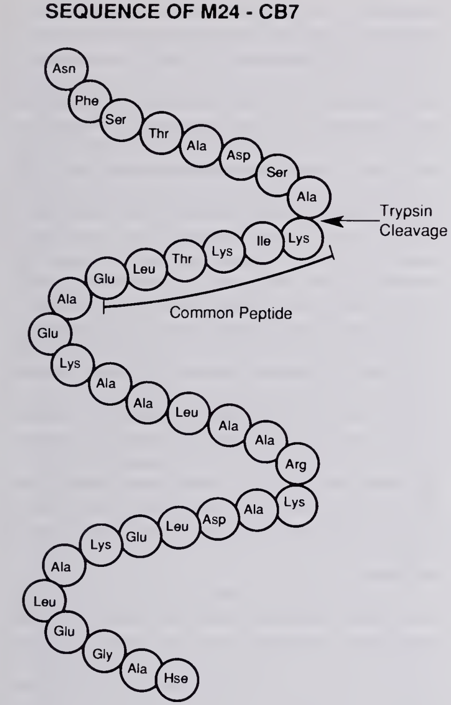

SEQUENCE OF M24 - CB7
Asn
Phe
Ser
Thr
Ala
Asp
Ser
Ala
Lys
Ile
Lys
Thr
Leu
Glu
Ala
Glu
Lys
Ala
Ala
Leu
Ala
Ala
Arg
Lys
Ala
Asp
Leu
Glu
Lys
Ala
Leu
Glu
Gly
Ala
Hse
Trypsin Cleavage
Common Peptide

Fig. 10. Purification of M24 protein: Amino acid sequence of a small amino-terminal peptide containing the type-specific epitope (76).

cross-reactive with heart tissues (88, 89), synovium (90), and brain (91) were located between the B and C repeats of the M5 molecule (Fig. 11). In addition, many of the cross-reactive epitopes were shared among several rheumatogenic M proteins, serotypes 5, 6, 18, and 19. All of these M proteins contain similar sequences within their B repeats which are the likely location of the shared cross-reactive tissue epitopes, heart, brain, and joint, that fit the spectrum of lesions of rheumatic fever.

Current M-protein Vaccine Strategies

By 1987, incorporation of several synthetic type-specific epitopes into a trivalent vaccine (M5, M6, and M24) was shown to be feasible (92). The trivalent vaccine evoked opsonic antibodies against all three serotypes and none of the sera cross-reacted with human tissues. By 1993, recombinant techniques produced more complex hybrid proteins containing greater numbers of M protein fragments that were also of greater length. The complimentary genes have been used to transform *Escherichia coli,* resulting in a recombinant tetravalent M protein immunogenic against four serotypes (93). Currently, the technique is being extended to an octovalent vaccine (94) **(Fig. 12).**

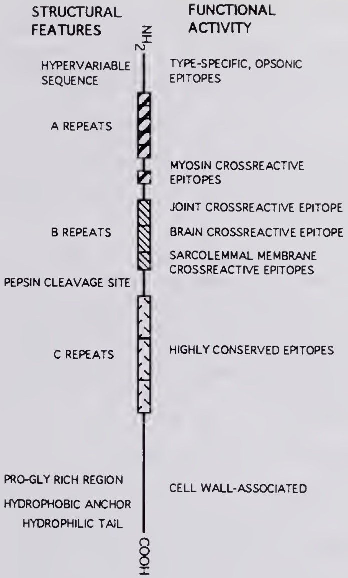

STRUCTURAL FEATURES
FUNCTIONAL ACTIVITY
NH2
HYPERVARIABLE SEQUENCE
TYPE-SPECIFIC, OPSONIC EPITOPES
A REPEATS
MYOSIN CROSSREACTIVE EPITOPES
JOINT CROSSREACTIVE EPITOPE
B REPEATS
BRAIN CROSSREACTIVE EPITOPE
SARCOLEMMAL MEMBRANE CROSSREACTIVE EPITOPES
PEPSIN CLEAVAGE SITE
C REPEATS
HIGHLY CONSERVED EPITOPES
PRO-GLY RICH REGION
HYDROPHOBIC ANCHOR
HYDROPHILIC TAIL
CELL WALL-ASSOCIATED
COOH

Fig. 11. Schema depicting the location on the M-protein molecule of the epitopes cross-reactive with myocardium, synovia, and brain. They are shown located between the B and C peptide repeats of the M5 molecule, well proximal to the type-specific epitopes on the amino-terminal pepM peptide (Ref. 94. Courtesy James B. Dale).

Additional strategies have included linking the recombinant hybrid proteins to the entire B subunit of *E. coli* labile toxin (LT-B), the latter serving not only as a haptenic carrier, but as a mucosal stimulant for the production of IgA type-specific antibodies **(Fig. 13).** As early as 1988, Bessen and Fischetti (95) showed the influence of nasal immunization with synthetic M peptides on mucosal colonization of GrAS (95), and Dale and

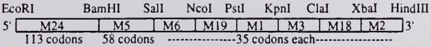

EcoRI
BamHI
SalI
NcoI
PstI
KpnI
ClaI
XbaI
HindIII
5'
M24
M5
M6
M19
M1
M3
M18
M2
3'
113 codons
58 codons
35 codons each

Fig. 12. Schema depicting a hybrid molecule synthesized by recombinant techniques to reproduce multiple type-specific M epitopes on a single immunogenic peptide (Ref. 93. Courtesy James B. Dale).

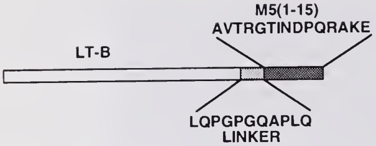


Fig. 13. Schema depicting the recombinant molecule in which the M5 epitope is linked to a haptenic carrier, the B subunit of *E. coli* labile toxin, LT-B (Ref. 96. Courtesy James B. Dale).

Chang (96) currently have demonstrated the presence of serum antibodies and salivary IgA in mice after intranasal immunization of their LT-B-5 vaccine (Fig. 13). Moreover, they were able to demonstrate protection in these mice after intraperitoneal challenge with virulent M5 streptococci. These observations are being extended to multivalent LT-B vaccines (94).

Future Directions for Vaccines

In addition to the limited number of serotypes associated with RF, the majority of life-threatening infections in this country are caused by a limited number of serotypes. Currently in the United States M1, M3, and M18 have been isolated most frequently from patients with serious invasive infections, and M3 and M18 from patients with RF. Among uncomplicated infections, the most common serotypes are 1, 2, 4, and 12. Multivalent vaccines might be devised based on disease and strain prevalence. The prevalence of nontypeable strains appears to vary depending on location (45). In the United States only 19% of the strains recovered from patients with uncomplicated infections were nontypeable, 26% from cases of severe infection were nontypeable, and none of the isolates from RF cases were nontypeable. In Thailand, however, only 20% of the GrAS isolates were typeable with available antisera, and a significant number were nontypeable in Australia (97). New serotypes in various regions will need to be identified. However, even in Northern Australia a significant percentage of strains associated with RF appear to belong to the M5 family of M proteins. It may be necessary to reformulate vaccine constructs periodically to represent prevalent serotypes of serious infections in various parts of the world.

Current Treatment of Streptococcal Pharyngitis in the United States

What should we do about diagnosing and treating the reservoir of streptococcal pharyngitis persisting in the United States? The common attribute of controlled studies designed for the prevention of RF was sustained penicillin treatment lasting at least 10 days. Where the prevalence of RF demanded the most stringent treatment, a single injection of 1.2 million units of benzathine penicillin G intramuscularly was considered first choice because compliance with regimens of oral penicillin G was unreliable (98). Nevertheless, oral regimens of 10 days of treatment with penicillin V, 2.0 g daily, in either four or, more effectively (99), two divided doses have become the norms, at least as far as "norms" can be determined for current treatment of sore throat (100).

Either throat culture or tests for GrAS antigen on throat swabs provide a powerful and useful negative predictive index by Baysian formulation, especially in settings in which RF is rare. In such settings, negative tests can therefore identify approximately 90% of sore throats as not due to GrAS and therefore not needing penicillin treatment. False-positive tests are commonly due to either prior infection or to the acquisition of throat-colonizing, noninvasive GrAS organisms. Treatment of these, however unnecessary, do little harm.

Follow-up throat cultures have not been advised when treatment has been adequate and pharyngeal symptoms have not relapsed. When such follow-up cultures have been made as part of clinical trails, however, some 6% to 29% of patients treated with recommended regimens continued to carry GrAS after clinical recovery (99). Such persistence of convalescent carriage has been overinterpreted as "failures" of therapy (100). In seminal studies in 1948, Rothbard and Watson (59) demonstrated the loss of M protein of such persisting strains after spontaneous recovery from infection which was correlated with progressive loss of virulence (see above). Thus, the effectiveness of penicillin does not depend on total eradication of GrAS (which cannot be accomplished). Success appears to depend on driving from the population the streptococcal strains that cause postinfectious complications, or in decreasing their virulence by preventing rapid person-to-person transmission (60). Where RF has disappeared so have the M-rich, mucoid (encapsulated) strains with which RF has been associated. The practical fact (confirmed wherever it has been studied) is that pharyngeal carriage persisting after streptococcal pharyngitis does not cause problems, provided the infection has been treated adequately with penicillin. The persistent organisms become attenuated and eventually disappear (60). High rates of failure to eradicate pharyngeal streptococci after standard penicillin therapy are suggestive of carriage of strains of low virulence in the cohorts reported. This observation is explained by the studies of Beachey and Ofek (72) on the avid adherence to the pharyngeal mucosa of strains that are devoid of M protein but retain the ligand, lipoteichoic acid.

Beta-lactamase-producing organisms in the throat (staphylocci or hemophilus species) have been associated with difficulty in eradicating GrAS (101), but this association does not appear to cause clinical problems and does not, in my view, justify eradicating these beta-lactamase producers with broad-spectrum antibiotics nor their use in the treatment of GrAS pharyngitis, particularly in view of the efforts now required to prevent emerging resistance (especially of *Pneumococci*) that is being fostered by the overzealous use of cephalosporins in relatively benign upper respiratory tract infections (102).

Better characterization of GrAS in study populations is needed, but, in populations where RF is still a major problem, injectable benzathine penicillin G has been brilliantly effective and is now an accepted treatment in countries still struggling with the problem. Until we mobilize resources for further intensive development of a vaccine against this wily organism, however, it will continue to use its apparently unlimited capacity for variation to survive in the human host and to cripple young hearts.

References

1. Rantz LA, Boisvert PJ, Spink WW. Etiology and pathogenesis of rheumatic fever. Arch Intern Med 1945; 76: 131–138.
2. Lancefield RC. Specific relationship of cell composition to biological activity of hemolytic streptococci. Harvey Lect 1941; 35:251.
3. Coburn AF. The factor of infection in the rheumatic state. Baltimore: Williams & Wilkens, 1931.
4. Todd EW. Antigenic streptococcal hemolysin. J Exp Med 1932; 55:267–280.
5. Coburn AF, Moore LV. The prophylactic use of sulfanilamide in streptococcal respiratory infections, with special reference to rheumatic fever. J Clin Invest 1939; 18:147–155.
6. Thomas CB, France R. A preliminary report of the prophylactic use of sulfanilamide in patients susceptible to rheumatic fever. Johns Hopkins Hosp Bull 1939; 64:67–77.
7. Rammelkamp CH, Denny FW, Wannamaker LW. Studies on the epidemiology of rheumatic fever in the armed services. In: Thomas L, ed. Rheumatic fever. Minneapolis: University of Minnesota Press, 1952: 72–89.
8. Stollerman GH, Rusoff JR. Prophylaxis against group A streptococcal infections in rheumatic fever patients. Use of a new repository penicillin. JAMA 1952; 150: 1571–1575.
9. Stollerman GH, Rusoff JH, Hirshfield I. Prophylaxis against group A streptococci in rheumatic fever. The use of single monthly injections of benzathine penicillin G. N Engl J Med 1955; 252:787–792.
10. Wood HF, Stollerman GH, Feinstein AR, et al. A controlled study of three methods of prophylaxis against streptococcal infection in a population of rheumatic children. N Engl J Med 1957; 106:345–356.
11. Chamovitz R, Catanzaro FJ, Stetson CA, et al. Prevention of rheumatic fever by treatment of previous streptococcal infections. 1. Evaluation of benzathine penicillin G. N Engl J Med 1954; 251:466–471.
12. Frank PF, Stollerman GH, Miller LF. Protection of a military population from rheumatic fever. JAMA 1965; 193:775–783.
13. Nelson JD, Mohs EA. Socioeconomic and behavioral residua of antibiotic therapy in infancy and childhood. In: Geraldoni Grassi G, Sabbath LD, eds. New trends in antibodies. New York: Elsevier/N. Holland Biomedical Press, 1981.
14. Taranta A, Stollerman GH. The relationship of Sydenham's chorea to infection with group A streptococci. Am J Med 1956; 20:170–175.
15. Taranta A. Relation of isolated recurrences of Sydenham's chorea to preceding streptococcal infections. N Engl J Med 1959; 260:1204–1210.
16. Johnson JC, Stollerman GH. Nephritogenic streptococci. Annu Rev Med 1969; 20:315–322.
17. Siegel AC, Johnson EE, Stollerman GH. Controlled studies of streptococcal pharyngitis in a pediatric population. 1. Factors related to the attack rate. N Engl J Med 1961; 265:559–566.
18. Stollerman GH. Rheumatic fever and streptococcal infection. New York: Grune and Stratton, 1975.
19. Stollerman GH, Siegel AC, Johnson EE. Variable epidemiology of streptococcal disease and the changing pattern of rheumatic fever. Mod Concepts Cardiovasc Dis 1965; 34:45–48.
20. Coburn AF, Pauli RH. Studies on the immune response of the rheumatic subject and its relationship to activity of the rheumatic process. IV. Characteristics of strains of hemolytic streptococcus, effective and noneffective in initiating rheumatic activity. J Clin Invest 1935; 14:755–762.
21. Kuttner AG, Krumwiede E. Observations on the effect of streptococcal upper respiratory infections on rheumatic children. A three-year study. J Clin Invest 1941; 20:273–287.
22. Rammelkamp CH Jr., Weaver RS, Dingle JH. Significance of the epidemiological differences between acute nephritis and acute rheumatic fever. Trans Assoc Am Physicians 1952; 65:168–175.
23. Bisno AL. The concept of rheumatogenic and non-rheumatogenic group A streptococci. In: Read SE, Zabriski JB, eds. Streptococcal diseases and the immune response. New York, NY: Academic Press, 1980: 789.
24. Bloomfield AL, Rantz LH. An outbreak of streptococci septic sore throat in an army camp. Clinical and epidemiologic observations. JAMA 1943; 121:315–319.
25. Dudding BA, Dillon HC, Wannamaker LW, et al. Postepidemic surveillance studies of a food-borne epidemic of streptococcal pharyngitis at the United States Air Force Academy. J Infect Dis 1969; 120:225–236.

26. Stollerman GH. Nephritogenic and rheumatogenic group A streptococci. J Infect Dis 1969; 120:258–263.
27. Parker MT, Basset DCJ, Maxted WR, et al. Acute glomerulonephritis in Trinidad: Serological typing of group A streptococci. J Hyg 1968; 66:657–675.
28. Top FH, Wannamaker LW, Maxted WR, et al. M antigens among group A streptococci isolated from skin lesions. J Exp Med 1967; 126:667–685.
29. Wannamaker LW. Medical progress. Differences between streptococcal infections of the throat and of the skin. N Engl J Med 1970; 282:23,78.
30. Potter EV, Svartman M, Mohammed I, et al. Tropical acute rheumatic fever and associated streptococcal infections compared with concurrent acute glomerulonephritis. J Pediatr 1978; 92:325.
31. Widdowson JP, Maxted WR, Grant DL. The production of opacity in serum by group A streptococci and its relationship with the presence of M antigen. J Gen Microbiol 1970; 61:343–353.
32. Widdowson JP, Maxted WR, Grant DL, et al. The relationship between M antigen and opacity factor in group A streptococci. J Gen Microbiol 1971; 65:69–80.
33. Widdowson JP, Maxted WR, Notley CM, et al. Antibody responses in man to infection with different serotypes of group A streptococci. J Med Microbiol 1974; 7:483.
34. Maxted WR, Valkenburg HA. Variation in the M-antigen of group A streptococci. J Med Microbiol 1969; 2:199–210.
35. Bisno AL, Pearce IA, Wall HP, et al. Contrasting epidemiology of acute rheumatic fever and acute glomerulonephritis. Nature of the antecedent streptococcal infection. N Engl J Med 1970; 283:561–565.
36. Bisno AL, Svartman M, Stollerman GH. Cross-reacting and monotypic T antigens of nephritogenic pyoderma streptococci: The type 56 system. J Immunol 1971; 106:1493–1498.
37. Land M, Bisno AL. Acute rheumatic fever: A vanishing disease in suburbia. JAMA 1983; 249:895.
38. Stollerman GH. Global changes in group A streptococcal diseases and strategies for their prevention. Adv Intern Med 1982; 27:373.
39. Holmberg SD, Faich FA. Streptococcal pharyngitis and acute rheumatic fever in Rhode Island. JAMA 1983; 250:2307–2312.
40. Stollerman GH. Rheumatogenic group A streptococci and the return of rheumatic fever. Adv Intern Med 1990; 35:1–26.
41. Veasy LG, Wiedmeier SE, Garth SO, et al. Resurgence of acute rheumatic fever in the intermountain area of the United States. N Engl J Med 1987; 316:421.
42. Centers for Disease Control. Acute rheumatic fever—Utah. MMWR 1987; 36:108.
43. Kaplan EL, Johnson DR, Cleary PP. Group A streptococcal serotypes isolated from patients and sibling contacts during the resurgence of rheumatic fever in the United States in the mid-1980s. J Infect Dis 1989; 159: 101–103.
44. Schwartz B, Facklan RR, Breiman RF. Changing epidemiology of group A streptococcal infection in the USA. Lancet 1990; 336:1167–1171.
45. Johnson DR, Stevens DL, Kaplan EL. Epidemiologic analysis of group A streptococcal serotypes associated with severe systemic infections, rheumatic fever or uncomplicated pharyngitis. J Infect Dis 1992; 166:374–382.
46. Kaplan EL. Global assessment of rheumatic fever and rheumatic heart disease at the close of the century. Circulation 1993; 88:1964–1972.
47. Cone LA, Woodard DR, Schlievert PM, et al. Clinical and bacteriologic observations of a toxic shock-like syndrome due to *Streptococcus pyogenes*. N Engl J Med 1987; 317:146.
48. Kim YB, Watson DW. Streptococcal exotoxins: Biological and pathological properties. In: Wannamaker LW, Matsen JM, eds. Streptococci and streptococcal diseases. New York: Academic Press, 1972: 33–50.
49. Stollerman GH. The historical role of the Dick test. JAMA 1983; 250:3097–3099.
50. Stollerman GH. Changing group A streptococci. The reappearance of streptococcal 'toxic shock.' Arch Intern Med 1988; 148:1268.
51. Stevens DL, Tanner MH, Winship J, et al. Severe group A streptococcal infections associated with toxic shock-like syndrome and scarlet fever toxin A. N Engl J Med 1989; 321:1–7.
52. Holm SE, Norrby A, Bergholm AM, Norgren M. Aspects of pathogenesis of serious group A streptococcal infections in Sweden, 1988–1989. J Infect Dis 1992; 166: 31–37.
53. Norrby-Teglund A, Newton D, Kotb M, Holm SE, Norgren M. Superantigenic properties of the group A streptococcal exotoxin SpeF (MF). Infect Immun 1994; 62:5227–5233.
54. Iwasaki M, Igarashi H, Hinuma Y, Yutsudo T. Cloning, characterization and overexpression of a *Streptococcus pyogenes* gene encoding a new type of mitogenic factor. FEBS Lett 1993; 331:187–192.
55. Schick B. Uber die Nachkrankheiten des Scharlach. Jahrb Kinderheik (Berlin) 1907; 65:132–175.
56. Hoge CW, Schwartz B, Talkington DF, et al. The changing epidemiology of invasive group A streptococcal infections and the consequence of streptococcal toxic shock-like syndrome. A retrospective population-based study. JAMA 1993; 269:384–389.
57. Bisno AL, Stevens DL. Streptococcal infections of skin and soft tissues. N Engl J Med 1995; 334:240–245.
58. Stollerman GH, Ekstedt RD. Long chain formation by strains of group A streptococci in the presence of homologous antiserum: A type-specific reaction. J Exp Med 1957; 106:345–356.
59. Rothbard S, Watson RF. Variation occurring in group A streptococci during human infection. Progressive loss of M substance correlated with increasing susceptibility to bacteriostasis. J Exp Med 1948; 87:521–533.
60. Krause RM, Rammelkamp CH Jr. Studies of the carrier stater following infection with group A streptococci. II. Infectivity of streptococci isolated during acute pharyngitis and during the carrier state. J Clin Invest 1962; 41:575–578.
61. Wilson AT. The relative importance of the capsule and the M antigen in determining colony form of group A streptococci. J Exp Med 1959; 109:257.
62. Bisno AL, Pearce IA, Stollerman GH. Streptococcal infections that fail to cause recurrences of rheumatic fever. J Infect Dis 1977; 136:278.
63. Berrios X, delCampo E, Guzman B, Bisno AL. Discontinuing rheumatic fever prophylaxis in selected adolescents and young adults: A prospective study. Ann Intern Med 1993; 118:401–406.

64. Fox EN, Pachman LM, Wittner MK, et al. Primary immunization of infants and children with group A streptococcal M protein. J Infect Dis 1969; 120:598–604.

65. Potter EV, Stollerman GH, Siegel AC. Recall of type specific antibodies in man by injections of streptococcal cell walls. J Clin Invest 1962; 41:301–310.

66. Stollerman GH. Prospects for a vaccine against group A streptococci. The problem of the immunology of M proteins. Arthritis Rheum 1967; 10:245–255.

67. Kaplan MH, Svec KH. Immunologic relation of streptococcal and tissue antigens. III. Presence in human sera of streptococcal antibody cross reactive with heart tissue. Association with streptococcal infection, rheumatic fever and glomerulonephritis. J Exp Med 1964; 119:651–666.

68. Zabriskie JB, Freimer EH. An immunological relationship between the group A streptococcus and mammalian muscle. J Exp Med 1966; 124:661–678.

69. Beachey EH, Stollerman GH. Mediation of cytotoxic effects of streptococcal M protein by nontype specific antibody in human sera. J Clin Invest 1973; 52:2563–2570.

70. Widdowson JP, Maxted WR, Pinney AM. An M-associated protein antigen (MAP) of group A streptococci. J Hyg (Camb) 1971; 69:553–564.

71. Vosti KL, Johnson RH, Dillon MF. Further characterization of purified fractions of M protein from a strain of group A type 12 streptococcus. J Immunol 1971; 107:104–114.

72. Beachey EH, Ofek I. Epithelial cell binding of group A streptococci by lipoteichoic acid on fimbriae denuded of M protein. J Exp Med 1976; 143:759.

73. Beachey EH, Campbell GL, Ofek I. Peptide digestion of streptococcal M protein. II. Extraction of M antigen from group A streptococci with pepsin. Infect Immun 1974; 9:891–896.

74. Swanson J, Hsu KC, Gotschlich EC. Electron microscopic studies on streptococci. I. M antigen. J Exp Med 1969; 130:1063.

75. Beachey EH, Stollerman GH, Chiang EY, et al. Purification and properties of M protein extracted from group A streptococci with pepsin. Covalent structure of the amino terminal region of type 24 M antigen. J Exp Med 1977; 145:1469.

76. Beachey EH, Stollerman GH, Johnson RH, Ofek I, Bisno AL. Human immune response to immunization with a structurally defined polypeptide fragment of streptococcal M protein. J Exp Med 1979; 150:862–877.

77. Dale JB, Beachey EH. Protective antigenic determinant of streptococcal M protein shared with sarcolemmal membrane protein of human heart. J Exp Med 1982; 156:1165–1176.

78. Dale JB, Beachey EH. Multiple heart-cross-reactive epitopes of streptococcal M proteins. J Exp Med 1985; 161:113–176.

79. Phillips GN, Flicker PF, Cohen C, et al. Streptococcal M protein: Alpha-helical coiled-coil structure and arrangements on the cell surface. Proc Natl Acad Sci USA 1981; 78:4689.

80. Manjula BN, Fischetti VA. Tropomyosin-like seven residue periodicity in three immunologically distinct streptococcal M proteins and its implications for the antiphagocytic property of the molecule. J Exp Med 1980; 151:695.

81. Hollingshead SK, Fischetti VA, Scott JR. Complete nucleotide sequence of type 6 M protein of the group A streptococcus. Repetitive structure and membrane anchor. J Biol Chem 1986; 261:1677–1686.

82. Bessen D, Jones KF, Fischetti VA. Evidence for two distinct classes of streptococcal M-protein and their relationship to rheumatic fever. J Exp Med 1989; 169:269–283.

83. Fischetti VA, Jones KF, Hollingshead SK, Scott JR. Structure, function and genetics of streptococcal M protein. Rev Infect Dis 1988; 10 (Suppl 2):S356–S359.

84. Dale JB, Seyer JM, Beachey EH. Type-specific immunogenicity of a chemically synthesized peptide fragment of type 5 streptococcal M protein. J Exp Med 1983; 158:1727–1732.

85. Beachey EH, Seyer JM. Protective and nonprotective epitopes of chemically synthesized peptides of the NH_2-terminal region of type 6 streptococcal M protein. J Immunol 1986; 136:2287–2292.

86. Beachey EH, Bronze MS, Dale JB, Kraus W, Poirier TP, Sargent SJ. Protective and autoimmune epitopes of streptococcal M protein vaccine. Vaccine 1988; 6:192–196.

87. Dale JB, Beachey EH. Sequence of myosin-cross-reactive epitopes of streptococcal M protein. J Exp Med 1986; 164:1785–1790.

88. Cunningham MW, McCormack JM, Fenderson PG, Ho MK, Beachey EH, Dale JB. Human and murine antibodies cross-reactive with streptococcal M protein and myosin recognize the sequence GLN-LYS-SER-LYS-GLN in M protein. J Immunol 1989; 143:2677.

89. Krisher K, Cunningham MW. Myosin: A link between streptococci and heart. Science 1985; 227:413–415.

90. Baird RW, Bronze MS, Kraus W, Hill HR, Veasey LG, Dale JB. Epitopes of group A streptococcal M protein shared with antigens of articular cartilage and synovium. J Immunol 1991; 146:3132–3137.

91. Bronze MS, Dale JB. Epitopes of streptococcal M proteins that evoke antibodies that cross-react with human brain. J Immunol 1993; 151:2820–2828.

92. Beachey EH, Seyer JM, Dale JB. Protective immunogenicity and T lymphocyte specificity of a trivalent hybrid peptide containing NH_2-terminal sequences of types 5, 6 and 24 M proteins synthesized in tandem. J Exp Med 1987; 166:647–656.

93. Dale JB, Chang EC, Lederer JW. Recombinant tetravalent group A streptococcal M protein vaccine. J Immunol 1993; 151:583–591.

94. Dale JB. Type-specific multivalent group A streptococcal M protein vaccines. In: Steven DL, Kaplan EL, eds. Streptococcal Infections. New York: Oxford University Press, 1996.

95. Bessen D, Fischetti VA. Influence of intranasal immunization with synthetic peptides corresponding to conserved epitopes of M protein on mucosal colonization by group A streptococci. Infect Immun 1988; 565: 2666–2672.

96. Dale JB, Chang EC. Intranasal immunization with recombinant group A streptococcal M protein fragment fused to the B subunit of *Escherichia coli* labile toxin protects mice against systemic challenge infections. J Infect Dis 1995; 171:1038–1041.

97. Hartas J, Goodfellow AM, Currie BS, Sriprakash KS. Characterization of group A streptococcal isolates from tropical Australia with high prevalence of rheu-

matic fever: Probing for signature sequences to identify members of the family of serotype 5. Microb Pathog 1995; 18:345–354.
98. Dajani AS, Bisno AL, Chung KJ, et al. Prevention of rheumatic fever: A statement for health professionals by the Committee on Rheumatic Fever, Endocarditis, and Kawasaki Disease of the Council on Cardiovascular Disease in the Young, the American Heart Association. Circulation 1988; 78:1082–1086.
99. Raz R, Elchanan G, Colodner R, et al. Penicillin V twice daily vs four times daily in the treatment of streptococcal pharyngitis. Infect Dis Clin Pract 1995; 4:50–54.
100. Stollerman GH. Penicillin for streptococcal pharyngitis: Has anything changed? Hosp Pract 1995; 30:80–83.
101. Picherero ME. Cephalosporins are superior to penicillin for the treatment of tonsillopharyngitis: Is the difference worth it? Pediatr Infect Dis J 1993; 123:268–274.
102. Stollerman GH. The global impact of penicillin: Then and now. Mt Sinai J Med 1993; 60(2):112–119.

The Emergence of "Emerging Diseases":

A Lesson in Holistic Epidemiology

Edwin D. Kilbourne, M.D.

Abstract

The term "emerging diseases" is a loosely defined category of entities comprising resurgent or recurrent old diseases (usually caused by "new" or mutated previously known agents), diseases truly new to man, but caused by preexisting ("old") zoonotic agents, and syndromes newly defined by the discovery of new agents through advances in biotechnology. Identification and solution of these problems depends, first, on recognition of their differences, and then upon tailoring appropriate strategies for their control. Thus, new influenza viruses appear each year to challenge immunity to their antecedents, but evoke the unchanged and centuries old symptom complex of influenza. Tuberculosis, is resurgent because of mycobacterial mutation to antibiotic resistance, immunosuppression by AIDS, and laxity in public health surveillance. Parvovirus B19 and herpesvirus 6 were revealed as cryptic infectors of white blood cells in studies of hepatitis B and AIDS, but since have been shown to be important causes of childhood rashes, aplastic anemia, and neurologic disease. The encroachment of human habitation on wilderness perimeters (ecosystem change) has increased contact with vectors of zoonotic viruses and bacteria, as evidenced by Lyme disease, Ebola virus infection, and the hemorrhagic fevers. The term "holistic epidemiology" embraces all these problems, from the molecular to the macroenvironmental level. Humans, parasites, and their environment will continue their ancient, fluctuating, dynamic relationship in the future, and new diseases will continue to emerge.

The persistence of ancient diseases into modern times has led us to assume a relative stability of host-parasite interaction and a constancy of disease expression. But today, just past the centenary of Louis Pasteur's death, and 200 years after Jenner's discovery of smallpox vaccine, we are shocked and concerned by the obstinacy of old infections and the emergence of new ones. Malaria long ago confounded the promise of dichlorodiphenyltrichloroethane (DDT) and tuberculosis is resurgent as a consequence of social disruption, resistance to chemotherapy, and the unforeseen destruction of host defenses by a virus new to recent human experience. This synergy of old and new influences on the pattern of disease illustrates the complexity of present problems and their resistance to simple solutions.

During the past few years, a state of complacency and an assumption that infectious diseases are under control in this antibiotic and vaccine age have been replaced by what I can only describe as the "emergence of emergent disease concern"—manifest, originally, by sober scientific consideration (1–7) but followed more recently by the "alarums and excursions" of full-scale media treatment.

The problem of new and emerging disease is a consequence not only of the mutability of parasites but of the interaction of changing microorganisms with changing human populations in dynamic and varying environments. The evolving industrial environment may yield nonmicrobial toxins no less threatening than the viruses of the jungle. The Spanish toxic oil syndrome and the American epidemic of eosinophilic myalgia from

---

A portion of this paper was presented at a Pasteur Centennial Conference on "Epidemiology and Public Health" in Hanoi, Vietnam on March 2, 1995.

From the Department of Microbiology and Immunology, New York Medical College. Address reprint requests to Edwin D. Kilbourne, M.D., Room 315, Basic Science Building, New York Medical College, Valhalla, NY 10595.

ingestion of a contaminated dietary supplement are notable examples. Dietary diseases of affluence and of poverty coexist within the same society.

Emerging diseases may or may not be nouvelle. To emerge is to "come into view." If the eye of the beholder is enhanced by modern technology, more diseases and agents of disease will appear as our threshold of detection is increasingly lowered. Diseases caused by parvoviruses and newly revealed human herpesviruses are not new, but recognition of these agents has opened the door to definition of their pathogenic effects. In the case of such obligate human pathogens, it is a fair assumption that they are ubiquitous, having evolved with man for centuries. Indeed, even agents that bear regional names (the spirochete of Lyme disease and the hantaviruses) are in fact widely distributed, as are their rodent carriers. The term "exotic viruses" seems no longer appropriate with its implications that "it can't happen here" and it is someone else's problem. There are agents, of course, that are constrained geographically, but these constraints exist only in balanced ecosystems. However, ecosystems are vulnerable to human intervention, as in the inadvertent and much publicized (8) importation of the African virus, Ebola, into the United States in infected monkeys.

New diseases may (occasionally) be new, but there is nothing new about the problem. *The problem, I believe, is to look soberly and without hysteria at exactly what the problem is.* When we do, it soon becomes apparent that "The Problem" is not one, but many separable problems, each of which must be isolated in traditional scientific fashion and evaluated separately. "The Problem," as addressed in many recent publications and symposia, would be better defined by what I should like to call "holistic epidemiology"—an emerging discipline, if you will, which comprises microbial and human genetics, human behavior, global ecology, toxicology, and environmental change. **Table 1** summarizes the disparate entities that comprise the loosely defined topic of "emerging diseases"—truly a potpourri.

Truly novel diseases are rare, and most are nonmicrobial. After the momentary triumph of the antibiotic age, we are faced with the reemergence or resurgence of old diseases such as tuberculosis, syphilis, streptococcal and pneumococcal disease which, if never really conquered, were coming under control in developed countries.

We seldom discuss the truly new diseases of this late 20th century whose existence has less to do with microbes than changing life-styles and the development of new technology. Some of these illnesses appear to be trivial, such as ocular and orthopedic problems related to the use of personal computers; on the other hand, some are not, as I shall describe later.

TABLE 1
Emerging Diseases—A Pot Pourri

1. Pristine (truly novel)
    - Nonmicrobial
    - Microbial
2. Reemerging old diseases
3. Nonmicrobial diseases
4. Diseases now increasing in prevalence
    - Old
    - New
5. Microbial diseases
    - Due to novel agents
    - Due to preexisting microbes newly emergent in man
    - Due to preexisting microbes newly recognized in man

The diseases now increasing in prevalence that are not necessarily emergent as new entities comprise both old and newly recognized diseases, some now linked together as are AIDS and tuberculosis. It is a paradox that mutated, variant, but nonetheless novel agents are most often found as causes of the old, well recognized disease syndromes. Other microbial diseases are due to "old" preexisting microbes suddenly transformed into pathogens as host defenses fail, and an increasing number of agents that are not new at all, but only newly recognized. With special emphasis on viral diseases, I should now like to consider, in reductionist fashion, the varied and multiple components of the "emerging disease" problem.

Microbial Evolution

Although the paradigm of emergent disease causation is the Andromeda Strain from outer space, most new, emergent, or reemergent diseases are not the consequence of extreme or bizarre mutations. Genetic evolutionary mechanisms are many (**Tables 2 and 3**), but so are constraints on microbial evolution (**Table 4**). This is not intended to minimize the role of mutation and evolution in the "restless tide" (1) of microbial diseases, but to put it in proper perspective. Indeed, the rapidity of influenza virus evolution through the acquisition of cumulative point mutations in its hemagglutinin antigen necessitates the development of new vaccines almost every year. In addition, rare instances of simultaneously mixed infections with human and animal influenza A viruses may lead to reassortment and

TABLE 2
Microbial Evolution

1. Cumulative point mutations
2. DNA recombination
    - Crossover
    - Cassette transfer
3. Genetic reassortment (segmented RNA viruses)
4. Bacterial gene transfer
    - Transformation
    - Transfection
    - Transduction
    - Conjugation
    - Plasmid transfer

TABLE 3
Facilitating Factors in Microbial Evolution

1. High mutation rates
2. Microbial genetic interchange
3. Immunosuppression of host permits emergence of minority mutants or infection by previously avirulent microbes
4. Host behavioral change may alter route or dose of infection
5. Ecologic changes in microbe-host balance

TABLE 4
Constraints on Microbial Evolution

1. Extreme mutations are lethal
2. "Cloning effect" during serial transfer
3. Critical virulence level must be maintained
4. Precise adaptation to an ecologic niche tolerates little genome variation
5. Major antigenic change requires multiple mutations
6. Zoonotic infections are rarely transmissible from man to man
7. Human MHC polymorphism limits emergence of "super-adapted" virus

Microbes constantly evolve but seldom "emerge" as new causes of disease.

recombination of their genes and the creation of truly novel pandemic viruses.

The RNA viruses, in general, have particularly high mutation rates, to the extent that they are recognized to comprise polymorphic subpopulations or "quasispecies" (9). This rate may exceed that of DNA-containing organisms by as much as 6 orders of magnitude. Like bacteria, viruses (both RNA and DNA) have the capacity to exchange genetic material and thus create essentially new agents of disease (as, for example, the influenza A viruses, above). Western equine encephalitis virus probably arose from the insertion of a "cassette" of genetic information from a Sindbis-like virus into the virus of eastern equine encephalitis virus (10).

Bacterial populations are particularly versatile in adapting to environmental changes within the host by the transfer of genes by a variety of mechanisms (Table 2). This genetic plasticity now confronts us with the evolution of tubercle bacilli, streptococci, and a host of other organisms that are no longer as vulnerable to antibiotics as they were only a few years ago.

The evolution of pathogenic microbes and viruses depends not only on their intrinsic mutation rates, but on other facilitating factors (Table 3). These include the capacity for genetic interchange with other microorganisms and physiologic and behavioral changes in the infected or target host that may contribute to the generation and selection of new mutants. Suppression of host immune response, whether by HIV infection or chemotherapy may 1) increase the magnitude and duration of viral replication, thereby increasing the probability of mutations, and 2) alter pathogenesis and expedite invasion of new tissues, adaptation to which site may involve the selection of virus with changed tissue specificity. Similarly, change in host behavior, e.g., intravenous drug administration, may deliver the pathogen (often in large quantity) to a novel site at which mutants may arise. Even though not selected immunologically, such mutants may be fortuitously changed antigenically (11).

It is now well established that changes in the microenvironment such as those induced by antibiotics in the gut can expedite and even foster the emergence to predominance of previously suppressed or inhibited species of bacteria, some of which may have pathogenic potential.

Constraints on Microbial Evolution

Viruses and microbes, particularly those of the latter category restricted to intracellular multiplication, must of necessity adapt to very specific ecologic niches to survive, to multiply, and to insure their own transmission and propagation (see Table 4). Thus, extreme mutations are lethal. In the case of the arboviruses that have both vertebrate and invertebrate hosts, adaptation to more than one ecologic niche is required.

Although a measure of virulence is necessary to ensure transmission of the infectious agent, extreme virulence is not to the parasite's advantage if the host is killed before transmission occurs. The maintenance in nature of certain infections, such as influenza, and probably other respiratory tract infections, is effected by transmission of very small doses of virus (12), thus, in effect, "cloning" the majority representative of the viral population with each serial passage. Such cloning diminishes the opportunity for the emergence of a minority mutant in the viral population. Viral genetic stability is maintained in part by the small number of particles required to transmit infection and the probability that they will be representative of the dominant subpopulation in the infecting host. However, persistent infections in cattle with plaque-cloned foot and mouth disease virus led to rapid selection of genetic and antigenic variants. There was sequential dominance of different viral subpopulations (13). Sequential evolution of mutants in the individual host has also been shown with experimental infection with swine influenza virus and during the course of natural infection with influenza virus in an immunocompromised child (14) and during infection with poliovirus type 2 (15).

In the laboratory, antigenically changed "escape" mutants can be selected readily by the pressure of immunoselection with monoclonal antibodies (16). However, epidemiologically significant antigenic changes usually require mutations at multiple sites to evade the polyclonal antibodies that are directed against more than one viral epitope.

Further constraints on unfettered microbial or viral evolution are 1) the barriers that exist against interhuman serial propagation of infection after interspecies transmission from animals, i.e., zoonoses, and 2) human major histocompatibility complex (MHC) polymorphism.

Disease Emergence Dependent On Microbial Change

Listed in **Table 5** is a mixed bag of various pandemic, panzootic, or resurgent diseases in which the common denominator is significant change in the causative parasite. Influenza appears to be unique among human infections in its requirement for continual and progressive antigenic change in its virus in nature for its survival. It is not yet clear that the equally frequent but not progressive antigenic variation within isolates of HIV plays a significant role in survival and transmission of that virus.

TABLE 5
Disease Emergence Dependent on Microbial Change

Influenza	
Point mutations	Epidemics
Gene reassortment	Pandemics
Cholera	(?) Pandemic
CTX toxin gene duplication	
Tuberculosis and many bacterial diseases	Resurgent
Antibiotic Resistance Genes	
AIDS	Epidemic/pandemic
Probable evolution from Simian ancestor	
Canine parvovirus disease	Panzootic
Three to four point mutations of feline parvovirus	
Canine distemper in lions	Epizootic

The notorious El Tor strain of *Vibrio cholerae,* which has been endemic in South East Asia for years, was introduced into Peru during the El Nino summer of 1991 by the discharge of bilge water from Asian harbors. (During the past decade it has been shown that the vibrio can survive for long periods within the algae of hot-weather algal blooms.) It has been speculated that mutation of the organism may be enhanced in the algal site through solar irradiation. In 1992, a new strain O139 arose, apparently in the Bay of Bengal, which had acquired multiple copies of the cholera toxin genetic element. It also differed antigenically from the less virulent El Tor. It has now spread from the confines of the Indian subcontinent and is considered to have pandemic potential (17).

The Price of Success (Sometimes) Is Failure

The successful treatment of tuberculosis with new drugs led to the rapid decline of infection, the closing of hospital wards and sanitoria, the shortening of treatment, the use of ambulatory and less well supervised treatment, the consequent development of antibiotic-resistant organisms, and the return of less tractable infection in a new, AIDS-immunosuppressed population in which the cycle has been amplified and the opportunity for further microbial evolution, enhanced.

Humans as Factors in the Emergence of Zoonoses—Viral Mutation, Alone, May Not Be Enough

Lions in the Serengeti may no longer be threatened by rifle bullets, but an estimated one third of the population has been wiped out recently by a variant of canine distemper virus. Domesticated dogs brought the epizootic to the perimeter of the preserve, and in apparently mutated form the virus rapidly spread to spotted hyenas, then to *Panthera leo* (18).

In the 1950s, raccoon rabies smoldered in Florida and Georgia and by the 1960s had reached South Carolina. In 1977, the disease suddenly jumped to West Virginia, and thence has spread rapidly in the Northeast. Finally crossing the Hudson, it is now at the New Hampshire border. Human intervention, in the form of hunters, was apparently responsible for the dislocation of the virus through the importation of raccoons as game animals from locales in which rabies had attained an enzootic state (19).

There are many genotypes of rabies virus, some of them apparently species-restricted as far as effective intraspecies transmission is concerned. When crossing species, rabies can kill with full virulence but will not spread within species subsequently, e.g., when fox bites raccoon. However, in the present case, rabies virus had become so exquisitely adapted to a raccoon population in one geographic area that it met with no resistance in northern raccoons in which there had been no prior selection for resistance. Thus, mixing of previously separated populations, even of the same species, can create epizootics or epidemics.

Disease Emergence Dependent on Changes in Host-Microbe Interaction (Table 6)

Increased Host-Microbe Density Increases Infection Probability And Microbial Dose (Social Determinants of CMV Infection)

Day care is used mostly by upper middle and middle class parents in whom childhood cytomegalovirus (CMV) infection had been less common. These parents are also more likely to breastfeed their infants. Their children in day care centers are now more likely to be infected early in life with CMV, to bring it home, and to infect their siblings and/or their mothers (20). Mothers of childbearing age may become infected and transmit the virus transplacentally to the unborn. However, one can see not only the prospect of increased congenital malformation from CMV, but the prospect of a decline in the epidemic in the next generation of progeny, who will be CMV-immune when they reach fertility and will not only resist primary infection during pregnancy but will pass along CMV virus in breast milk (and probably colostral antibodies), thus restoring the probable ancient balance of host and parasite that was seen, until recently, in the poorer population. The poorer population, on the other hand, increasingly does not breastfeed and does not use day care; therefore, they may experience a reverse in their epidemic pattern as their children, deprived of the benefits of early infection, will become like the well-to-do of today, vulnerable to primary infection during pregnancy.

TABLE 6

Disease Emergence Dependent on Changes in Host-Microbe Interaction

Increased host-microbe density	
Increases infection probability	
Increases microbial dose	Child care centers
Altered infection patterns	
Postchildhood contact	Polio, maternal CMV
Altered modes of infection	
Vaginal hygiene	Staphylococcal toxic shock syndrome
Organ transplantation	Various herpesviruses
Intravenous drug use	Hepatitis viruses

Like crowded military barracks of the past, child day care centers seem to function as microbial breeders, fostering the transmission of both bacterial and viral diarrheal diseases. For children, this poses an increased risk (and possibly severity) of infection, and for adults in contact with these children, primary acquisition of infection at a less desirable time of life.

Altered Infection Patterns

The incidence, prevalence, and severity of disease can be influenced without intrinsic changes in host or parasite but by changes in social patterns as described above for CMV. Any factors which postpone primary infection until adulthood increase the risk of severe disease at that later age. Indeed, the emergence of paralytic poliomyelitis in the early twentieth century was the result of improvements in environmental sanitation which postponed acquisition of infection until older childhood or later. Subsequently, the emergence of a generation uninfected by polioviruses in childhood because of successful vaccination created a population of young adults, some of whom were vulnerable to the effects of wild or even vaccine polioviruses excreted by their own children.

Altered Modes of Infection

An unfortunate by-product of medical progress has been the iatrogenic transmission of latent viruses during the course of organ transplantation, a problem which promises to become even more serious with the introduction of so-called xeno (cross species) organ transplantation.

"Progress" in feminine hygiene by the use of vaginal tampons led to an epidemic a decade ago of staphylococcal toxic shock syndrome.

Of Mice and Men—Environmental Change And the Appearance of New Diseases

Disruption in agricultural ecosystems and redistribution of human populations are now well known to be associated with increased human-rodent contact in the genesis of hemorrhagic fevers. We have seen the emergence of the new syndrome of Acute Respiratory Distress Syndrome in the far West, a hantavirus infection acquired from deer mice (21). Earlier, the burgeoning cities in Southeast Asia that encroached on rice paddies had resulted in transmission of other hantaviruses to humans from rodent sources.

Closer to home, in New York City, rickettsialpox has reappeared in a city hospital (22). Reminiscent of bubonic plague, mites (or fleas) will leave the bodies of their rodent hosts (for those of humans) either when the mice die or when their physiology or body temperature is disturbed. In the case of rickettsialpox, lymphocytic choriomeningitis infection (of mice) led to this result. Thus, although the proximate cause of rickettsialpox in man is the Rickettsia, the precipitating event may be a virus infection of mice, also dangerous to man, or the visitation of the exterminator!

The Sins of Our Fathers—The Origin Of Lyme Disease in New England

In the forests primeval, a balanced ecosystem existed with large stands of hardwood, grazing animals, and their predators. Reminiscent of today's Oregon woodcutters, farmers converted woodland to pasture, thereby eliminating not only the hardwood but the deer, mice, and the predators sheltered therein. The past two centuries have seen a decline in New England agriculture and the regrowth of woodland and the return of deer and mice, but not their predators, hunted to extinction 100 to 200 years ago. The return of the deer (if not the mice) was cheerfully welcomed as they invaded essentially scrub forests bordering on increasing suburban sprawl. With the deer and mice came the mite and with the mite came Borrelia and Lyme disease—one more argument for the balanced ecosystem (reviewed in Reference 23).

Virulence and Transmissibility May Be Linked—The Case of Dengue

In the case of dengue fever, the urban opportunism of an "inefficient vector," i.e., *Aedes egypti*, as a domesticated, rather indolent mosquito, has resulted in the selection of a vector that requires a high viremia to transmit infection; thus it spreads only the most virulent strains. Added to this are an increasingly cluttered urban environment (old tires, cans, etc.) and one has epidemic disease in Puerto Rico. Add to this the potential for the introduction of new viral variants to reinfect and the "new" syndrome of hemorrhagic fever emerges (reviewed in Reference 24).

Human Genetic Determinants of Disease Severity—Measles in the New World and the Cost Of Consanguinity

The common diseases of childhood in Europe and Asia behaved like exotic agents of high virulence in American Indians and South Sea Islanders. Francis Black at Yale has spent years looking for genetically based immunologic defects in these populations without success (25). Their response to live virus vaccines is normal. They have no obvious defects in cell-mediated immunity or IgG response to antigens. However, they are quite homogeneous (and restricted) with respect to MHC polymorphism.

Each of the host's MHC antigens presents a different set of viral peptides to the immune system. This leads to a selection of specialized viral progeny less likely to be recognized by the MHC-limited host population so that a virus of greater virulence for that population emerges. Black calculated that there is a 32% chance that a virus passing between two Amerinds will not encounter a new MHC type at either the A or B locus. Ordinarily, in an outbred population, a secondary host is likely to have a different set of MHC genes than the index case. Therefore, different viral peptides are presented to the immune system in that instance and a variable viral consensus population will be maintained.

Modern studies of measles in Senegal by Garenne and Aaby (26) documented increasing disease severity after passage of virus between close relatives compared with parallel passage of virus in unrelated persons. This phenomenon was attributed to a dose effect as the virus moved from casual village contacts into family compounds in which contact and exposure was increasingly intensified. The risk of dying increased 1.6-fold each generation of disease. Black believed that this is better explained based on his hypothesis, i.e., that selection of more virulent virus has occurred with serial replication in a consanguineous population. The answer may be, both.

Revelation of Common Etiology Creates "New" Diseases by Linkage of Disparate Manifestations (Table 7)

The Rashes of Childhood

Two cryptic, but very different viruses have been recognized recently as inhabitants of human blood cells:

Parvovirus B19 is a small DNA virus whose canine relative, carried on human shoes from continent to continent, caused pandemic disease in domestic and feral canines (27). Since its initial discovery, during the course of serologic investigations of hepatitis B virus by electron microscopy in which characteristic parvovirus particles were found in immunoprecipitates, the virus has been linked to erythema infectiosum ("slapped cheek syndrome") (5th Disease) and later to juvenile and occasionally adult symmetrical polyarthritis (28). It has also been implicated in aplastic anemia crises and as a cause of neonatal death in hydrops fetalis. It has been found in persist in clotting factor concentrates.

Human herpesvirus 6 (29) was uncovered as a by-product of HIV investigations. Although it has been found in normal peripheral lymphocytes, it is (like parvovirus B19) a transmissible contagious agent of human disease (roseola infantum) and is ubiquitous in the population. It has been linked to exanthem subitum in childhood, but may have a more serious role in infection of marrow transplant recipients and in central nervous system infections.

A lesson here is that the signs and symptoms of expression of infection as disease are sometimes poor indicators of the nature of and pathogenic potential of the infectious agent. Clinical surveillance alone sometimes is a poor guide to the characterization of new diseases. Definitive definition of the etiologic agent, microbial or not, can link disparate symptoms in the recognition of "new" or newly recognized syndromes, as discussed further below.

TABLE 7
Revelation of Common Etiology Creates "New" Diseases by Linkage of Disparate Manifestations

- Human herpesvirus 6
    - Roseola infantum
    - Mononucleosis
    - Central nervous system disease
- Parvovirus B19
    - Erythema infectiosum
    - Juvenile polyarthritis
    - Aplastic anemia crises
    - Hydrops fetalis
- Toxic cooking oil, dietary tryptophan
    - Eosinophilia myalgia syndrome

TABLE 8
Conclusions

Dramatic encounters with exotic zoonotic pathogens has galvanized medical (and public) attention.

Simultaneously, containment of traditional human disease is failing.

New diseases (old agents) and old diseases (new agents), reemerging and resurgent diseases and diseases simply increasing in prevalence have been combined as a single, large, confusing problem.

These separable problems should be addressed separately.

The term "holistic epidemiology" embraces all these problems from the molecular to the macroenvironmental level.

Man, parasite, and environment will continue their ancient, fluctuating, dynamic relationship in the future.

Nonmicrobial Toxicity, Dietary Nonsense, and the Eosinophilia-Myalgia Syndrome

In the spring and summer of 1981, an epidemic of a new illness affected 20,000 people in central and northwestern Spain, resulting in 12,000 hospitalizations and more than 300 deaths in the first year of the epidemic. Patients had acute onset of a respiratory syndrome characterized by cough, dyspnea, hypoxemia, pulmonary infiltrates, and pleural effusion. Fifty percent progressed to a chronic phase of myalgia, eosinophilia, and peripheral nerve damage. Painstaking investigation on a somewhat cold trail linked the epidemic to ingestion of cooking oil adulterated with rapeseed oil denatured with aniline (30).

In October 1989, three patients in New Mexico presented with myalgia, weakness, skin rash, and striking eosinophilia. All patients were using L-tryptophan as a dietary supplement. Centers for Disease Control investigation with alerts to all states revealed more than 1500 cases of the disease, all ultimately linked to a single Japanese manufacturer and a contaminant, ethylidenebistryptophan (31).

Discussion and Conclusions

The present decade has seen the culmination of a series of quite independent forces and events which have been uncritically categorized together as "emerging diseases," with the implication that they have a common genesis and therefore a common solution. The terrifying emergence of Ebola virus from the African jungles has arrested public attention without appreciation of the silent but far more important resurgence of such "conquered" diseases as tuberculosis and the continuing, yearly toll of influenza 50 years after the development of a vaccine. Paradoxically, genomically new (mutated, recombined, or plasmid-bearing) agents are most often implicated in recurrences of these old diseases, whereas preexisting microbes (but heretofore alien to or usually nonpathogenic for man) account for most new clinical syndromes. The aging of the world's population, the continuing spread of AIDS, and the increasing use of immunosuppressive chemotherapy for neoplasia and organ transplantation all predispose to unusual manifestations of common infections. All these effects are compounded by environmental and social changes that affect emergence of disease by increasing the probability of host-parasite contact by altered vector proximity or by presentation by an abnormal route of exposure.

It is unlikely that present trends favoring the emergence of new diseases will soon be reversed. Furthermore, the price of current advances in medical biotechnology (gene therapy, DNA vaccines, organ transplantation) is the introduction of new potential hazards. We can no longer rely on the semiempirical development of new generations of antibiotics; rather, there must be a return to the principles of basic science in the design of future drugs and vaccines and in the study of human behavior that may lead us into harm's way.

References

1. Krause RM. The restless tide: The persistent challenge of the microbial world. The National Foundation for Infectious Diseases, 1981.
2. Kilbourne ED. Are new diseases really new? Nat Hist 1983; 92:28–32.
3. Lederberg J, Shope RE, Oaks SC, eds. Emerging infections: Microbial threats to health in the United States. Washington, DC: National Academy Press, 1992.
4. Morse SS, Schluederberg A. Emerging viruses: The evolution of viruses and viral diseases. J Infect Dis 1990; 162:1–7.
5. Kilbourne ED. New viral diseases: A real and potential problem without boundaries. JAMA 1990; 264:68–70.
6. Morse SS, ed. Emerging viruses. New York: Oxford University Press, 1993.
7. Murphy FA. New, emerging and reemerging infectious diseases. Adv Virus Res 1994; 43:1–52.
8. Kilbourne ED. Book review. Nat Med 1995; 1:89–90.
9. Holland J, Spindler K, Horodyski F, Grabau E, Nichol S, Vandepol S. Rapid evolution of RNA genomes. Science 1982; 215:1577–1585.
10. Hahn CS, Lustig S, Strauss ED, Strauss JH. Western equine encephalitis is a recombinant virus. Proc Natl Acad Sci USA 1988; 85:5997–6001.
11. Kilbourne ED. Genetic dimorphism in influenza viruses: Characterization of stably associated hemagglutinin mutants differing in antigenicity and biological properties. Proc Natl Acad Sci USA 1978; 75:6258–6262.
12. Alford RH, Kasel JA, Gerone PJ, et al. Human influenza resulting from aerosol inhalation. Proc Soc Exp Biol Med 1966; 122:800–804.
13. Gebauer F, de la Torre JC, Gomes I, et al. Rapid selection of genetic and antigenic variants of foot-and-mouth disease virus during persistence in cattle. J Virol 1988; 62:2041–2049.
14. Rocha E, Cox NJ, Black RA, Harmon MW, Harrison CJ, Kendal AP. Antigenic and genetic variation in influenza virus isolates recovered from a persistently infected immunodeficient child. J Virol 1991; 65:2340–2350.
15. Kinunnen I, Huovilainen A, Poeyry T, Hovi T. Rapid molecular evolution of wild type poliovirus during infection in individual hosts. J Gen Virol 1990; 71:317–324.
16. Wiley DC, Wilson IA, Skehel JJ. Structural identification of the antibody-binding sites of Hong Kong influenza hemagglutinin and their involvement in antigenic variation. Nature 1981; 289:373–378.
17. Waldor MK, Mekalanos JJ. Emergence of a new cholera pandemic: Molecular analysis of virulence determinants in *Vibrio cholerae* 0139 and the development of a live vaccine prototype. J Infect Dis 1994; 170:278–283.
18. Roelke-Parke ME, Munson L, Packer C, et al. A canine distemper virus epidemic in Serengeti lions (*Panthera leo*). Nature 1996; 379:441–445.
19. Krebs JW, Strine TW, Smith JS, Rupprecht CE, Childs JE. Rabies surveillance in the United States during 1993. J Am Vet Med Assoc 1994; 205:1695–1709.
20. Stagno S, Cloud GA. Working parents: The impact of day care and breast feeding on cytomegalovirus infections in offspring. Proc Natl Acad Sci USA 1994; 91:2384–2389.
21. Hughes JM, Peters CJ, Cohen ML, Mahy BWJ. Hantavirus pulmonary syndrome: An emerging infectious disease. Science 1993; 262:850–851.
22. Kass EM, Szaniawski WK, Levy H, Leach J, Srinivasan K, Rives C. Rickettsialpox in a New York City hospital, 1980 to 1989. N Engl J Med 1994; 331:1612–1617.
23. Spielman A. In: Englund PT, Sher A, eds. The biology of parasitism. New York: Alan R. Liss, Inc., 1988.
24. Halstead SB. Global epidemiology of dengue hemorrhagic fever. Southeast Asian J Trop Med Public Health 1990; 21:636–637.
25. Black F. An explanation of high death rates among New World peoples when in contact with Old World diseases. Perspect Biol Med 1994; 37:292–303.
26. Garenne M, Aaby P. Pattern of exposure and measles mortality in Senegal. J Infect Dis 1991; 161:1088–1094.
27. Parrish CR, Have P, Foreyt WJ, Evermann JF, Senda M, Carmichael LE. The global spread and replacement of parvovirus strains. J Gen Virol 1988; 69:1111–1116.
28. Cohen B. Parvovirus B19: An expanding spectrum of disease. Br Med J 1995; 311:1549–1552.
29. Lusso P, Gallo RC. Human herpesvirus 6. Ballieres Clin Haematol 1995; 8:201–223.
30. Kilbourne EM, Posaado de la Paz M, Abaitus Borda I, Diez Ruiz-Navarro M, Philen RM, Falk H. Toxic oil syndrome: A current clinical and epidemiologic summary. J Am Coll Cardiol 1991; 18:711–717.
31. Kilbourne EM. Eosinophilia-myalgia syndrome: Coming to grips with a new illness. Epidemiol Rev 1992; 14:16–36.

Research in Psychodynamic Psychotherapy

Robert S. Wallerstein, M.D.

Introduction: Outcome and Process

The central research questions in psychoanalysis and in the expressive and supportive psychoanalytic psychotherapies qua therapies are: 1) What changes take place during and as a consequence of the therapy (i.e., the outcome question), and 2) how do those changes come about or how are they brought about—that is, through the interaction of what factors in the patient, in the therapist and treatment, and in the evolving life situation (i.e., the process question). In theory, process and outcome are necessarily interlocked. Any study of outcome, even if it counts only the percentage of cases "improved," must establish some criteria for "improvement," and these criteria, in turn, must derive from some conceptualization of the nature of the illness and the process of change, whether or not such a conceptualization is explicitly formulated. Similarly, any study of process, in delineating patterns of change among variables, makes cross-sectional assessments at varying points in time, which, if compared with one another, provide measures of short- or longer-term outcomes.

On practical grounds, however, outcome studies and process studies are usually separated. Although process is conceptually not separable from outcome, methods that yield the best judgments in the one area are often operationally opposed to those that yield the best judgments in the other. For example, judgments of outcome will be scientifically most convincing if bias is minimized and freedom from contamination maintained by keeping those who make the "after" judgments unaware of the "before" judgments and predictions. From the point of view of process judgments about the same therapy, such care to minimize contamination would be unnecessary. Indeed, it would be counter to the whole spirt of inquiry into process, in which maximum knowledge of all the known determinants, as these have varied over time, is essential to best understand the changes that occur.

However, in view of the space constraints of this paper, and the greater importance placed on outcome studies, or efficacy studies as they are often called, from the point of view of public policy and the deployment of fiscal resources within the healthcare delivery system, I will focus here on psychotherapy outcome research and refer the reader interested in concomitant process-focused studies to the expanded version of this article in Wallerstein (1). I will also focus here on formal and systematic empirical research rather than on the far larger and much antecedent clinical case study and theoretical literature out of which, of course, have grown the questions to which the more formal research studies have been addressed. And finally, in this statement of my frame, I should mention some of the considerations related to the topic of outcome (and also process) research in psychoanalytic therapies—issues that, although beyond the scope of this arti-

---

This article was published, in expanded form, in *Psychodynamic Concepts in General Psychiatry,* edited by Harvey J. Schwartz with Efrain Bleiberg and Sidney H. Weissman, American Psychiatric Press, 1995.

Professor Emeritus and former Chairman, Department of Psychiatry, University of California San Francisco School of Medicine, and Training and Supervising Analyst, San Francisco Psychoanalytic Institute. Also Past President, American Psychoanalytic Association, 1971–2, and Past President, International Psychoanalytical Association, 1981–5.

Address correspondence to Robert S. Wallerstein, M.D., 655 Redwood Highway, Suite 261, Mill Valley, CA 94941.

cle, are nonetheless germane to its central argument. These include:

1. The goals of these treatment modalities, both ideal and practical (realizable).
2. The issues of suitability or treatability, as against analyzability, which is not the same thing, although the two are often confounded.
3. The indications and contraindications for this array of treatments as these have evolved over time with increasing experience and expanding theoretical and technical knowledge.
4. The role of the initial diagnostic and evaluation procedures in differential treatment planning (as against the view that only a trial of treatment can lead to proper case formulation and prognostication).
5. The place of prediction (and predictability) in relation to issues of change, expectable reach, and limitation.
6. The theory of technique (how treatment works and achieves its goals; that is, the relationship of means to ends).
7. The similarities and differences between psychoanalysis and the dynamic psychotherapies, as compared from the viewpoint of different goals projected for different kinds of patients, determining through these differences the specifically appropriate technical approaches from within the available spectrum.
8. The criteria for satisfactory treatment termination.
9. The evaluation of results (process and outcome changes, therapeutic benefit, and analytic completeness).
10. The conception of the ideal state of mental health and the unavoidable impingements upon its empirical assessment by value judgments as well as by the vantage point and the partisan interests of the judge.
11. The place of the follow-up study as a desirable, feasible, and appropriate activity (or not) in relation to psychoanalytic therapies, for research and/or for clinical purposes.
12. The place of the continuing accretion of knowledge in relation to all these areas by the traditional case study method innovated by Freud as against the desirability or necessity for more formal systematic clinical research into these issues, by methods, of course, that are responsive to the subtlety and complexity of the subjectivistic clinical phenomena, while simultaneously remaining loyal to the canons of empirical science.

Early Statistical Studies: First-Generation Outcome Research

As early as 1917, within the first decades of the introduction of psychoanalysis to America, Coriat (2) reported on the therapeutic results achieved in 93 cases, of which 73% were declared either recovered or much improved, these rates being nearly equal across all his diagnostic categories. As with all the early statistical studies to be noted here, the judgments of improvement were made by the treating clinician, according to (usually) unspecified criteria, and without individual clinical detail or supporting evidence.

In the 1930s, several comparable but larger scale reports emerged from the collective experiences of the psychoanalytic treatment centers of some of the pioneering psychoanalytic training institutes. Fenichel (3) reported results from the initial decade of the Berlin Institute, the first formally organized psychoanalytic institute in the world. Of 1,955 consultations conducted, 721 patients were accepted for psychoanalysis; 60% of the neurotic patients were judged to have received substantial benefit, but only 23% of those who were adjudged "psychotic."[1] Jones (4) reported on 738 applicants to the London Psychoanalytic Clinic, of whom 74 were taken into analysis; 28 of the 59 neurotic patients were judged to have benefited substantially, but only 1 of the 15 so-called psychotic patients. Alexander (5) reported on 157 cases from the Chicago Psychoanalytic Clinic, with 63% of the neurotic patients, 40% of the psychotic patients, and 77% of those designated psychosomatic judged to have received substantial benefit from their treatment. During the same period, Kessel and Hyman (6), two internists who followed up 29 patients they referred for psychoanalysis, reported that almost all the neurotic patients had benefited from the treatment, and all the psychotic patients were either unchanged or worse.

In a review article evaluating the overall results of psychoanalysis, Knight (7) combined the findings of all of these studies (except the first one by Coriat) and added 100 patients treated at the Menninger Clinic, where the results were judged to be completely comparable to those of the other studies in the observed outcomes with neurotic and psychotic patients. The overall composite tabulation comprised 952 patients, with the therapeutic benefit rate approximately 60% for the neurotic, close to 80% for the psychosomatic, and only 25% for the psychotic patients. Knight made particular reference to the pitfalls of these simple statistical summaries: the absence of consensually agreed upon definitions and criteria, the crudity of nomenclature and diagnostic classification, and the failure to address issues of therapeutic skill in relation to cases of varying severity.

---

[1] Psychotic here is to be taken as ambulatory, and in some sense functioning in the community, although considered "psychotic" from the standpoint of the quality of mental life.

The most ambitious study of this first-generation genre was the report of the Central Fact-Gathering Committee of the American Psychoanalytic Association (8).[2] Data were collected during a 5-year period beginning in 1952; altogether, there were 10,000 initial responses to questionnaires submitted by 800 analysts and candidates, with some 3,000 termination questionnaires submitted upon completion of treatment. As with the other studies cited thus far, criteria for diagnosis and improvement were unspecified, and these and other flaws and ambiguities resulted ultimately in a report that was simply an "experience survey" consisting of 1) facts about the demographics of analysts' practices, 2) analysts' opinions on their patients' diagnoses, and 3) analysts' opinions on the therapeutic results achieved. Not unexpectedly, most patients were declared substantially improved.

Finally, in 1968, Feldman (9) reported on the results of psychoanalysis in 120 patients (selected from 960 evaluations) treated at the clinic of the Southern California Psychoanalytic Institute over its 11-year history. Again, the reported improvement rates were completely comparable to those of all the preceding studies, with two thirds of the outcomes being in the "good" or "very good" categories; and once more, difficulties were experienced due to the lack of clear and agreed upon criteria, concepts, and language for diagnostic assessment, analyzability, and analytic results.

Altogether, this sequence of so-called first-generation outcome studies of psychoanalysis, spanning a half-century from 1917 to 1968, was scientifically simplistic and failed to command the interest of the clinical psychoanalytic world. Most practitioners agreed with Glover (10) in his dour assessment: "Like most psycho-therapists, the psychoanalyst is a reluctant and inexpert statistician" (p. 393)—and, we could add, researcher. Such conclusions spurred the second-generation studies—the efforts at more formal and systematic outcome research geared toward overcoming the glaring methodologic simplicity that marked each of the studies described to this point.

Formal and Systematic Outcome Studies: Second-Generation Research

The methodologic flaws in the first-generation statistical enumerations of psychoanalytic outcomes have already been indicated. In addition to the lack of consensually agreed upon criteria at almost every step, from initial assessment to outcome judgments of therapeutic benefit and analytic result, and the use of these judgments (derived from unspecified, and even unformulated, criteria) by the (necessarily biased) therapist, as the (usually) sole evidential primary database, there was also the methodologic difficulty that these studies were all retrospective, with all the potential therein for bias, confounding and contamination of judgments, *post hoc ergo propter hoc* reasoning, and the like. Efforts to address these issues, including the introduction of prospective inquiry and even the fashioning of predictions to be validated or refuted by subsequent assessment, began in earnest in the 1950s and 1960s. Three major projects based on studies of clinic cases from the Boston, Columbia, and New York Psychoanalytic Institutes are described from this second-generation research approach.

Group-Aggregated Studies

Boston Institute. In 1960, Knapp et al. (11) reported on 100 supervised psychoanalytic patients from the Boston Institute clinic, rated initially for suitability for analysis, of whom 27 were followed up a year later with questionnaires addressed to the treating analysts to ascertain just how suitable the patients had turned out to be. The evaluation procedures (i.e., the initial judgments and the subsequent questionnaire responses) were "blind." There turned out to be fair but limited success in this assessment of suitability for analysis from the initial evaluation. However, two significant limitations of this study should be mentioned. First, the testing of the predictions took place only at the 1-year mark rather than more suitably upon treatment termination; clearly, much can change in this regard, in both directions, at later points in analysis. Second, and this is a problem with all research on this model, the patients selected by psychoanalytic clinic committees are already carefully screened, with obviously unsuitable cases already rejected. The range of variability in the accepted cases is thus considerably narrowed, making differential prediction within that group inherently less reliable.

---

[2] Although the span from Coriat (2) to this study (8) is a half-century, I call all of these studies "first generation" in terms of their degree of conceptual and methodologic sophistication rather than in temporal terms—although, of course, each "generation" started at a later point in time than its predecessor, or spanned a later period.

Sashin et al. (12), subsequently studied 183 patients treated at the same clinic from 1959 to 1965. Final data were collected on 130 (72%) of these patients after an average of 675 treatment hours and at a point averaging 6 years after treatment termination. Predictor variables were assessed with a 103-item evaluation questionnaire and via six major outcome criteria: 1) restriction of functioning by symptoms, 2) subjective discomfort, 3) work productivity, 4) sexual adjustment, 5) interpersonal relations, and 6) availability of insight. Only 10 of the predictor items demonstrated some predictive value in relation to assessed outcomes and that only with modest (albeit statistically significant) correlation. As a group, however, these predictor variables "made little clinical sense." Overall, the Boston Institute studies yielded only fair prediction to judgments of analyzability as assessed at the 1-year mark in treatment, and no effective prediction at all to treatment outcomes from the patient's characteristics as judged at initial evaluation.

Columbia Psychoanalytic Center. The Columbia Psychoanalytic Center project, contemporaneous with the Boston studies, was written up in a final accounting in a series of published reports in 1985 (13–16). This project consisted of prospective studies of a large number of patients (1,348 in sample 1, and 237 in the later sample 2), all treated by the same group of therapists. Data were collected from multiple perspectives over time (initially and at termination), with opportunities to compare findings in psychoanalysis (about 40% of the total) with those in psychoanalytic psychotherapy (the other 60%). The authors stated that all previous studies had been limited in at least one of the following ways: small sample size, inadequate range of information about outcomes, not based on terminated cases, or restricted to retrospective data. Further, no other studies had permitted comparison between large numbers of terminated analyses and psychotherapies conducted by the same analysts, in which all pertinent information had not been retrospectively assessed. In addition, criteria for therapeutic-benefit were established distinct from separate criteria for the evolution of a psychoanalytic process.

The most striking finding from this project was that across every category of patient, the therapeutic benefit measures always substantially exceeded the measures of an evolved analytic process. For example, only 40% of those who had completed analyses with good therapeutic benefit were characterized as having been "analyzed" by the project criteria. An equally striking finding was that the outcomes of these treatments, in terms of both therapeutic benefit and analyzability, were only marginally predictable from the perspective of the initial evaluation. This finding was in keeping with those of the Boston studies just cited, of course, and for the same reasons. As noted by Weber et al. (14) "The prudent conclusion . . . is not that therapeutic benefit or analyzability are per se unpredictable, but that once a case has been carefully selected as suitable for analysis by a candidate, its eventual fate remains relatively indeterminate" (p. 135).

Another significant finding was that "retrospective assessments of patient qualities by the treating analyst show a more substantial relationship to outcome than assessments made at the beginning of treatment" (14, p. 136). As expected, those selected for psychoanalysis were assessed initially as functioning at higher levels than those selected for psychotherapy, and those in psychoanalysis achieved greater therapeutic benefits than those in psychotherapy. In conclusion, the authors stated that sample 1 was three times larger than that in any previously published study, and that the project was the first to have a psychotherapy comparison group and one of the first to make the conceptual distinction between analyzability and therapeutic benefit. Sample 2 was a smaller sample gathered a decade later with some refinements in methods of data collection and some differences in observational vantage points, but in almost every particular all the findings of sample 1 were replicated. From both studies, the authors concluded, "that a substantially greater proportion of analysands derive therapeutic benefit than develop an analytic process, and that the development of an analytic process is associated with the highest levels of therapeutic benefit. Yet, what we do not yet know precisely is the nature and quality of benefit associated with the development of an analytic process and without its development" (15, p. 261).

The final article in this series (13) was devoted to a review of clinical and methodologic considerations. The authors stressed the advantages of their project over other comparable studies: 1) the N was very large; 2) it was a prospective study with predictive evaluations performed before outcomes were known; 3) they used many (clinically meaningful) scales; 4) aside from evaluations by patients and therapists, they used independent judges; and 5) psychoanalysis and psychotherapy were assessed comparatively.

New York Psychoanalytic Institute (Erle and Goldberg). The New York Psychoanalytic Institute studies (17–19) were similarly constituted, although with greater focus on the study of treatments carried out by more experienced analysts. There were two studies. The first (17) consisted of 40 supervised analytic patients selected from 870 applicants to the Treatment Center of the New York Psychoanalytic Institute. The results were completely comparable with those of the Boston and Columbia centers. Twenty-five of the patients terminated satisfactorily but only 11 of those were considered to have completed treatment; 24 of the patients were judged to have benefited substantially, but only 17 were judged to have been involved in a proper psychoanalytic process. A sample of 42 private patients from seven analytic colleagues of the author who began treatment in the same period and were assessed in the same manner as the Treatment Center patients showed substantially comparable results. The second study (19) extended the work to a sample of 160 private patients gathered over a subsequent 5-year span from 16 cooperating experienced analysts. The outcomes from these experienced analysts were completely comparable to the results of their own (and of all other) earlier studies of clinic patients treated by candidates.

Individually Focused Studies

New York Psychoanalytic Institute (Pfeffer). Over a time span parallel with these relatively large-sample outcome studies of psychoanalytic clinic patient populations (as well as some comparison private patients) assessed by pre- and/or post-treatment rating scales and grouped statistically, Pfeffer initiated a completely different type of outcome and follow-up study of terminated psychoanalyses by intensive individual case studies of a research-procured population at the same New York Psychoanalytic Institute Treatment Center (20–22). He first reported on nine patients who had completed analyses under Treatment Center auspices and had agreed to a series of follow-up interviews by a "follow-up analyst" who had not conducted the treatment. The interviews were open-ended, once a week, "analytic" in character, and ranged from two to seven in number before the participants agreed upon a natural close. The chief finding, in *all* instances, consisted of the rapid reactivation of characteristic analytic transferences. This included even transitory symptom flare-ups, as if in relation to the original treating analyst, with subsequent rapid subsidence, at times aided by pertinent interpretations, and in a manner that indicated the new ways of neurotic conflict management that had been achieved in the analysis.

In the last of these reports (22), Pfeffer attempted a metapsychologic explanation of these "follow-up study transference phenomena" (p. 230). He concluded that, "The recurrence in the follow-up study of the major preanalytic symptomatology in the context of a revived transfence neurosis as well as the quick subsidence of symptoms appear to support the idea that conflicts underlying symptoms are not actually shattered or obliterated by analysis but rather are only better mastered with new and more adequate solutions" (p. 234). The neurotic conflicts thus "lose their poignancy" (p. 237).

San Francisco and Chicago Research Groups. Two other research groups, one in San Francisco (23,24) and one in Chicago (25–27), replicated the Pfeffer studies, with some slight alterations in method, and in both instances confirmed what has come to be called "the Pfeffer phenomenon." The San Francisco group concluded that, "the infantile neurosis had not disappeared. What had changed was the degree to which it affected [the patient's] everyday life" (23, p. 492). The Chicago group concluded that, "psychic conflicts were not eliminated in the analytic process . . . the more significant outcome of the analysis appeared to be the development of a preconsciously active self-analytic function, in identification with the analyzing function of the analyst, as a learned mode of coping with conflicts" (27, p. 9). This overall finding from all three groups—that even in analyses considered highly successful, neurotic conflicts are not obliterated or shattered, as was once felt, but rather are tamed, muted, or lose their poignancy—is echoed in the well-known analytic quip that we all still recognize our good friends after their analyses.

I note one final consideration: A shared characteristic of these second-generation studies—whether the group-aggregated broad statistical accountings (the Boston, the Columbia, Erle and Goldberg in New York) or the individually focused, in-depth research studies (Pfeffer in New York, the San Francisco, the Chicago)—was the failure to segregate outcome results discerned at treatment termination from the issue of the stability (or not) of these results as revealed at some established follow-up point subsequent to termination, with all the different possibilities—for consolidation and further enhancement of treatment gains, for the simple maintenance of treatment achievements, or for actual regression back toward the pretreatment state.

Conceptually, this was a failure to accord specific theoretical status to what Rangell (28) has called "the postanalytic phase." Rangell described a variety of possible courses that can characterize this phase and concluded that "the desired goal should be a transition to a normal interchange in which the analyst can be seen and reacted to as a normal figure and no longer as an object for continued transference displacement" (28, p. 722). In the third-generation studies next to be described, the distinction between results at the termination study point and those at a subsequent prearranged follow-up study point (anywhere from 2 to 5 years later) becomes a clearly demarcated research focus, among the advances over the second-generation studies. A considerable variety of post-termination therapist-patient contact or interaction patterns are delineated in these studies.

Combined Process and Outcome Studies: Third-Generation Research

What I am calling the third-generation studies of the outcomes of psychoanalysis have actually been contemporaneous in time with the (conceptually) second-generation studies just described. These are systematic and formal psychoanalytic therapy research projects that have attempted both to assess analytic outcomes across a significant array of cases and to examine the processes through which these outcomes have been reached via the intensive longitudinal study of each of the individual cases. In this, these studies have combined the methodologic approaches of the group-aggregated studies (the Boston, the Columbia, and the New York, Erle and Goldberg) with those of the individually focused studies (the New York, Pfeffer, the San Francisco, and the Chicago). Like the best of the second-generation studies, they have carefully defined their terms, constructed rating scales, and tried to operationalize their criteria at each assessment point. These third-generation studies have been constructed prospectively, starting with pretreatment assessment of patients. Unlike the second-generation studies, they have carefully separated outcomes at termination from functioning at a specified subsequent follow-up point and have attempted to account for the further changes, in either direction, that took place during this "postanalytic phase." Bachrach and colleagues (29), in a comprehensive survey of research on the efficacy of psychoanalysis, singled out the newer Boston Psychoanalytic Society and Institute studies (30–37) and the Psychotherapy Research Project of The Menninger Foundation (38, 39) as the only ones to fully meet their array of state-of-the art specifications.

Boston Psychoanalytic Society and Institute Studies

The Boston studies were undertaken in the 1970s and came to publication in the following decade. Twenty-two supervised analytic cases at the Boston Institute Clinic were selected for prospective study, with the initial assessment based on a projective test battery used to yield measures of 1) affect availability and modulation, 2) quality of object relations, 3) adequacy of reality testing, and 4) motivation for change. Approximately a year after treatment termination, the initial test battery was repeated; also both the patient and the treating analyst were interviewed.

A series of three papers (31–33) described the results. Nine of the 22 patients had a successful analytic outcome, 5 had a limited outcome, and 8 were unanalyzed. Nonetheless, the greater number achieved therapeutic benefits along each of the change and outcome dimensions, and along each dimension the therapeutic benefit exceeded the analytic result in terms of the degree of successfully completed analytic work. A consistent and important finding was that therapeutic benefit was achieved by the majority of the patients and was regularly in excess of what could be accounted for by the evocation and the interpretive resolution of the transference neurosis. However, although most patients derived significant therapeutic benefit from their analytic experience, successful outcome could not be predicted from any of the predictor variables. This finding led these investigators to speculate that, "a particularly important omission (from the predictor variables) might have been consideration of the effect of the [therapist-patient] match in shaping the two-person psychoanalytic interaction" (34, p. 899). By "match" they meant, "an interactional concept: it refers to a spectrum of compatibility and incompatibility of the patient and analyst which is relevant to the analytic work" (34, p. 894). They further noted that although, "this mesh of the analyst's personal qualities with those of the patient has rarely been a special focus of attention . . . most analysts when making referrals do consider it; few assume that equally well-trained analysts are completely interchangeable" (30, p. 273).

This same team then returned for follow-up interviews with the same patient cohort as in 1987, now 5 to 10 years after the treatment terminations, this time including the retrospective assessment of the goodness of the analyst-patient match as one of the variables contributing to the patient outcomes (35–37). Nineteen of the original 22 patients could be located and of these, 18 agreed to be interviewed. Again a variety of change measures were used: global improvement ratings, affect management, quality of object relations, adequacy of reality testing, work accomplishment, and overall self-esteem. Overall results at the follow-up point comprised three patients further improved, four stable, six deteriorated but restored with additional treatment, four deteriorated despite additional treatment, and one returned to the original analyst and still in treatment and therefore not counted.

The most striking finding, however, was that the stability of achieved gains in the follow-up period could not be predicted from the assessments at termination—that is, according to Kantrowitz and colleagues (35), "psychological changes were no more stable over time for the group of patients assessed as having achieved a successful analytic outcome concomitant with considerable therapeutic benefit than for the other group of patients assessed as having achieved therapeutic benefit alone" (p. 493). In focusing on the assessment of analyst-patient match (37), the authors concluded that with 12 of the 17 patients, the kind of match (impeding or facilitating) did play a role in the outcome achieved. They gave examples of facilitating matches with good outcomes, impeding matches with poor outcomes, and more complex situations in which the kind of match was at first facilitating to the unfolding of the analytic process but later seemed to have an influence in preventing the completion of the analytic work.

Psychotherapy Research Project (The Menninger Foundation)

The other so-called third-generation psychoanalytic therapy research study to be described here is the Psychotherapy Research Project (PRP) of The Menninger Foundation, the most comprehensive and ambitious such research program ever carried out (38–40). The intent of the PRP was to follow the treatment careers and subsequent life careers of a cohort of patients (ultimately 42 in number), half in psychoanalysis, and half in other psychoanalytic psychotherapies—and each in the treatment deemed clinically indicated—from the initial pretreatment comprehensive psychiatric evaluation, through the entire natural span of their treatments (for however many years that entailed), . . . then into formal follow-up inquiries at several years after the treatment terminations, and with as much of an open-ended follow-up thereafter as circumstances might make possible and as the span of interested observation might last. The patients entered into their treatment during the mid-1950s (contemporaneous with the bulk of the second-generation studies); their periods of treatment ranged from 6 months to 12 years; all (100%) were reached for formal follow-up study at the 2- to 3-year mark; and more than one third could be followed up for periods ranging from 12 to 24 years beyond their treatment terminations (with four still in ongoing treatment in the early 1980s, 30 years after the inception of the project).

The aim of the PRP was to learn as much as possible about 1) what changes actually take place in psychoanalysis and other psychoanalytic psychotherapies (the outcome question), and 2) how those changes come about—through the interactions over time of what variables in the patient, in the therapy and the therapist, and in the patient's evolving life situation that co-determine those changes (the process question). Three overall treatment groups were set up, psychoanalysis, expressive psychoanalytic psychotherapy, and supportive psychoanalytic psychotherapy, in accordance with the consensus at that time in the psychoanalytic literature regarding the defining characteristics of these therapeutic modes, together with differential indications derived from the dynamic formulations of the nature of the patients' lives, history, character, and illness structure.

The project goals within this framework were to specify in detail the particular reach and limitation of the therapeutic outcome for each kind of patient appropriately treated within each of the proffered therapeutic approaches. There was a special interest in the more empirical elaboration of the psychologic change mechanisms operative within both the uncovering (expressive) and the "ego-strengthening" (supportive) therapeutic modes. *Forty-Two Lives in Treatment,* written from 1981 through 1982, represents the full statement of the project's findings and conclusions (38). For an overall capsule summarization of the main highlights here, I can best report by paraphrasing (and sharply condensing) a lengthy segment from the concluding part of a summarizing paper about the project (39, pp. 144–149).

PRP Conclusions. The overall conclusions can be combined as a series of sequential propositions regarding the appropriateness, efficacy, reach, and limitations of psychoanalysis and of psychoanalytic psychotherapies (variously expressive and supportive)—always, of course, with the caveat that this segment of the overall patient population (i.e., those [usually sicker] individuals who were brought to or who sought their intensive analytic treatment within a psychoanalytic sanatorium setting) was not necessarily representative of the usual outpatient psychoanalytic therapy population.

1. The first proposition concerns the distinction so regularly made between true "structural change," presumably based on the interpretive resolution of unconscious intrapsychic conflicts, and "behavioral change," based on "just altered techniques of adjustment" and presumably all that can come out of the other, noninterpretive, non-insight-aiming change mechanisms. Intrinsic to this way of dichotomizing change has always been the easy assumption that only structural change, as brought about through conflict resolution marked by appropriately achieved insight, can have some guarantee of inherent stability, durability, and capacity to weather at least ordinary future environmental vicissitude. The PRP experience clearly and strongly questioned the continued usefulness of this effort to so tightly link the kind of changes achieved with the intervention mode—expressive or supportive—by which they are effected. The changes reached in the more supportive therapies seemed often enough to represent just as much structural change, in the terms indicated, as the changes reached in the most expressive-analytic cases.
2. The second proposition concerns the conventional proportionality argument—that therapeutic change will be at least proportional to the degree of achieved conflict resolution. This proposition is, of course, almost unexceptionable, because it is clear that there can be significantly more change than there is true conflict resolution, on all the varying (supportive) bases through which change can be brought about, as well as properly proportionate change in which the change is on the basis of conflict resolution with accompanying insight, if such an ideal type ever actually exists in practice. Conversely, it would be hard to imagine real conflict resolution (and accompanying insight) without at least proportional concomitant change in behaviors, dispositions, and symptoms.
3. The third proposition, often linked to the proportionality argument but much more debatable and clearly separated from it, concerns the necessity argument, that effective conflict resolution is a necessary condition for at least certain kinds of change. It is clear that an overall PRP finding, and almost an overriding one, has been the repeated demonstration that a substantial range of changes—in symptoms, in personality functioning, and in life-style—have been brought about via the more supportive psychotherapeutic modes, cutting across the gamut of declared supportive and expressive (even analytic) therapies, and that in terms of the usual criteria, these changes can be (in many instances) quite indistinguishable from the changes brought about by typically expressive-analytic means.
4. A counterpart to the proposition based on the tendency to overestimate the necessity of the expressive (analytic) treatment mode's operating via conflict resolution to effect therapeutically desired change has been the other proposition, based on the happy finding that the supportive therapeutic approaches so often achieved far more than was expected of them—and, indeed, often enough reached the kinds of changes expected to depend on more expressive and insightful conflict resolutions—and did so in ways that represented indistinguishably "structural" changes. In fact, proportionately, each within its own category, the designated psychotherapy cases did as well as the designated psychoanalytic ones. More to the point, the good results in the one modality were not overall less stable or less enduring or less proof against subsequent environmental vicissitude than in the other. And, more important still, within both the psychotherapy and the psychoanalysis groups, the changes predicted, although more often predicated on the more expressive techniques, in fact were achieved to a greater-than-expected degree (often the same changes) on the basis of the more supportive techniques.
5. Considering these PRP treatments from the point of view of psychoanalysis as a treatment modality, just as more was accomplished than expected with psychotherapy, especially in its more supportive modes, so psychoanalysis, as

the quintessentially expressive therapeutic mode, achieved less, at least with these patients, than had been anticipated or predicted. This more limited success in part reflected the ethos of the psychoanalytic sanatorium and the psychoanalytic treatment opportunities this setting is intended to make possible; that is, the protection and life management of those temporarily disorganized and incompetent individuals who cannot be helped sufficiently with any other or lesser treatment approach than psychoanalysis, but who cannot tolerate the rigors of psychoanalytic treatment within the usual out-patient setting.

This is the concept of the so-called heroic indications for psychoanalysis, and in the PRP experience, these particular patients characteristically did very poorly with the psychoanalytic treatment method, however modified by parameters or buttressed by concomitant hospitalization. And there were certainly enough instances of very good outcomes with these "sicker" patients in an appropriately modulated supportive-expressive psychotherapy, especially when we could ensure truly adequate concomitant life management. The shift is in the departure from the effort at psychoanalysis per se (even modified analysis) as the treatment of choice for these "sicker" patients in that setting. On this basis, we can speak of the failing of the so-called heroic indications for psychoanalysis and can invite a repositioning of the pendulum, in its swings over time around this issue, more in the direction of "narrowing indications" for (proper) psychoanalysis.

6. The predictions made for prospective courses and outcomes tended to be for more substantial and more enduring change (more "structural change") where the treatment was to be more expressive-analytic; *pari passu,* the more supportive the treatment was intended to be (had to be), the more limited and inherently unstable the anticipated changes were expected to be. All of this speculation was consistently tempered and altered in the actual implementation in the treatment courses. Psychoanalysis and the expressive psychotherapies as a whole were systematically modified to introduce more supportive components, and they by and large accomplished less than anticipated, with a varying but often substantial amount of their successful outcomes achieved by supportive means. The supportive psychotherapies, on the other hand, often accomplished a fair amount more, and sometimes a great deal more, than initially expected, again, with much of the change based on more supportive techniques than originally specified.

In conclusion, these overall PRP results can be generalized as follows: 1) the treatment results in psychoanalysis and in varying mixes of expressive-supportive psychotherapies tend to converge, rather than diverge, in outcome; 2) across the whole spectrum of treatment courses, from the most analytic-expressive to the most single-mindedly supportive, the treatment carried more supportive elements than originally projected, and these elements accounted for substantially more of the changes achieved than had been originally anticipated; 3) the supportive aspects of psychotherapy, as conceptualized within a psychoanalytic theoretical framework, deserve far more respectful specification than they have usually been accorded in the psychodynamic literature; and 4) from the study of the kinds of changes achieved by this patient cohort, partly on an uncovering insight-aiming basis and partly on the basis of the opposed covering-up varieties of supportive techniques, the changes themselves, divorced from how they were effected, often seemed quite indistinguishable from each other in terms of being so-called real or structural changes in personality functioning.

Conclusion

In this article I have presented a review of the major formal and systematic outcome research studies (or efficacy studies) of psychoanalysis and related psychoanalytic psychotherapies. I have divided these outcome studies into three "generations" characterized by increasing complexity and sophistication, from first-generation early statistical studies (simple enumerations), through second-generation more formal and systematic outcome studies, and into third-generation combined process and outcome studies. Space considerations have precluded my reviewing the parallel development and the contemporary burgeoning of psychoanalytic therapy process research studies; interested readers are referred to *Psychoanalytic Process Research Strategies,* edited by Dahl et al. (41) I have also not reviewed so-called "offset" research relating mental health to general health care, based on the demonstration that the costs of intensive outpatient psychoanalytic therapy can in many instances be significantly offset by savings in medical expenditures for general health care including reduced use of hospitalization. Again, interested readers are referred to the early German studies by Duehrssen and Jorswick (42) and the more recent American studies by Mumford et al. (43, 44) and Schlesinger et al. (45, 46).

References

1. Wallerstein RS. Research in psychodynamic therapy. In: Schwartz HJ, Bleiberg E, Weissman SH, eds. Psychodynamic concepts in general psychiatry. Washington, DC and London: American Psychiatric Press, 1995: 431–456.
2. Coriat, I. Some statistical results of the psychoanalytical treatment of the neuroses. Psychoanal Rev 1917; 4:209–216.
3. Fenichel O. Statistischer bericht uber die therapeutische Tatigkeit 1920–1930. In: Zehn jahre Berliner Psychoanalytisches Institut. Internazionale Psychoanalytischer Verlag 1930:13–19.
4. Jones E. Decannual report of the London Clinic of Psychoanalysis, 1926–1936. London: London Clinic of Psychoanalysis, 1936.
5. Alexander F. Five year report of the Chicago Institute for Psychoanalysis: 1932–1937. Chicago, IL: Chicago Institute for Psychoanalysis, 1937.
6. Kessel L, Hyman H. The value of psychoanalysis as a therapeutic procedure. JAMA. 1933; 101:1612–1615.
7. Knight RP. Evaluation of the results of psychoanalytic therapy. Am J Psychiatry 1941; 98:434–446.
8. Hamburg DA, Bibring, GL, Fisher C, et al. Report of ad hoc committee on Central Fact-Gathering Data of the American Psychoanalytic Association. J Am Psychoanal Assoc 1967; 15:841–861.
9. Feldman F. Results of psychoanalysis in clinic case assignments. J Am Psychoanal Assoc 1968; 16:274–300.
10. Glover E. The indications for psycho-analysis. J Ment Sci 1954; 100:393–401.
11. Knapp PH, Levin S, McCarter RH, Wermer H, Zetzel E. Suitability for psychoanalysis: A review of one hundred supervised analytic cases. Psychoanal Q 1960; 29:459–477.
12. Sashin JI, Eldred SH, van Amerongen ST. A search for predictive factors in institute supervised cases: A retrospective study of 183 cases from 1959 to 1966 at the Boston Psychoanalytic Society and Institute. Int J Psychoanal 1975; 56:343–359.
13. Bachrach HM, Weber JJ, Solomon M. Factors associated with the outcome of psychoanalysis (clinical and methodological considerations): Report of the Columbia Psychoanalytic Center Research Project (IV). Int Rev Psychoanal 1985; 12:379–388.
14. Weber JJ, Bachrach HM, Solomon M. Factors associated with the outcome of psychoanalysis: Report of the Columbia Psychoanalytic Center Research Project (II). Int Rev Psychoanal 1985; 12:127–141.
15. Weber JJ, Bachrach HM, Solomon M. Factors associated with the outcome of psychoanalysis: Report of the Columbia Psychoanalytic Center Research Project (III). Int Rev Psychoanal 1985; 12:251–262.
16. Weber JJ, Solomon M, Bachrach HM. Characteristics of psychoanalytic clinic patients: Report of the columbia Psychoanalytic Center Research Project (I). Int Rev Psychoanal 1985; 12:13–26.
17. Erle JB. An approach to the study of analyzability and analysis: The course of forty consecutive cases selected for supervised analysis. Psychoanal Q 1979; 48:198–228.
18. Erle JB, Goldberg DA. Problems in the assessment of analyzability. Psychoanal Q. 1979; 48:48–84.
19. Erle JB, Goldberg DA. Observations on assessment of analyzability by experienced analysts. J Am Psychoanal Assoc 1984; 32:715–737.
20. Pfeffer AZ. A procedure for evaluating the results of psychoanalysis: A preliminary report. J Am Psychoanal Assoc 1959; 7:418–444.
21. Pfeffer AZ. Follow-up study of a satisfactory analysis. J Am Psychoanal Assoc 1961; 9:698–718.
22. Pfeffer AZ. The meaning of the analyst after analysis: A contribution to the theory of therapeutic results. J Am Psychoanal Assoc 1963; 11:229–244.
23. Norman HF, Blacker KH, Oremland JD, Barrett WG. The fate of the transference neurosis after termination of a satisfactory analysis. J Am Psychoanal Assoc 1976; 24: 471–498.
24. Oremland JD, Blacker KH, Norman HF. Incompleteness in "successful" psychoanalysis: A follow-up study. J Am Psychoanal Assoc 1975; 23:819–844.
25. Schlessinger N, Robbins FP. Assessment and follow-up in psychoanalysis. J Am Psychoanal Assoc 1974; 22:542–567.
26. Schlessinger N, Robbins FP. The psychoanalytic process: Recurrent patterns of conflict and changes in ego functions. J Am Psychoanal Assoc 1975; 23:761–782.
27. Schlessinger N, Robbins FP. A developmental view of the psychoanalytic process: Follow-up studies and their consequences. New York: International University Press, 1983.
28. Rangell L. An overview of the ending of an analysis. In: Litman RE, ed. Psychoanalysis in the Americas. New York: International University Press, 1966:141–165. Reprinted in Rangell L. The human core: The intrapsychic basis of behavior. vol. 2. Madison, CT: International University Press, 1990:703–725.
29. Bachrach HM, Galatzer-Levy, R, Skolnikoff AZ, Waldron S Jr. On the efficacy of psychoanalysis. J Am Psychoanal Assoc 1991; 39:871–916.
30. Kantrowitz JL. The role of the patient-analyst "match" in the outcome of psychoanalysis. Ann Psychoanal 1986; 14:273–297.
31. Kantrowitz JL, Paolitto F, Sashin J, Solomon L, Katz AL. Affect availability, tolerance, complexity, and modulation in psychoanalysis: Follow-up of a longitudinal, prospective study. J Am Psychoanal Assoc 1986; 34:529–559.
32. Kantrowitz JL, Katz AL, Paolitto F, Sashin J, Solomon L. Changes in the level and quality of object relations in psychoanalysis: Follow-up of a longitudinal prospective study. J Am Psychoanal Assoc 1987; 35:23–46.
33. Kantrowitz JL, Katz AL, Paolitto F, Sashin J, Solomon L. The role of reality testing in psychoanalysis: Follow-up of 22 cases. J Am Psychoanal Assoc 1987; 35:367–385.
34. Kantrowitz JL, Katz AL, Greenman D, et al. The patient-analyst match and the outcome of psychoanalysis: A pilot study. J Am Psychoanal Assoc 1989; 37:893–919.
35. Kantrowitz JL, Katz AL, Paolitto F. Follow-up of psychoanalysis five to ten years after termination. I: Stability of change. J Am Psychoanal Assoc 1990; 38:471–476.
36. Kantrowitz JL, Katz AL, Paolitto F. Follow-up of psychoanalysis five to ten years after termination. II: Development of the self-analytic function. J Am Psychoanal Assoc 1990; 38:637–654.
37. Kantrowitz JL, Katz AL, Paolitto F. Follow-up of psychoanalysis five to ten years after termination. III: The relation between resolution of the transference and the

patient-analyst match. J Am Psychoanal Assoc 1990; 38:655–678.

38. Wallerstein RS. Forty-two lives in treatment: A study of psychoanalysis and psychotherapy. New York: Guilford Press, 1986.
39. Wallerstein RS. Psychoanalysis and psychotherapy: Relative roles reconsidered. Ann Psychoanal 1988; 16:129–151.
40. Wallerstein RS, Robbins LL, Sargent HD, Luborsky L. The Psychotherapy Research Project of The Menninger Foundation: Rationale, method and sample use. Bull Menninger Clin 1956; 20:221–278.
41. Dahl H, Kachele H, Thoma H, eds. Psychoanalytic Process Research Strategies. Berlin: Springer-Verlag, 1988.
42. Duehrssen A, Jorswick E. An empirical and statistical inquiry into the therapeutic potential of psychoanalytic treatment. Der Nervenarzt 1965; 36:166–169.
43. Mumford E, Schlesinger HJ, Glass GV. The effects of psychological intervention on recovery from surgery and heart attacks: An analysis of the literature. Am J Public Health 1982; 72:141–151.
44. Mumford E, Schlesinger HJ, Glass GV, et al. A new look at evidence about reduced cost of medical utilization following mental health treatment. Am J Psychiatry 1984; 141:1145–1158.
45. Schlesinger HJ, Mumford E, Glass GV. Effects of psychological intervention on recovery from surgery. In: Guerra F, Aldrete JA, eds. Emotional and psychological responses to anesthesia and surgery. New York: Grune & Stratton, 1980:9–18.
46. Schlesinger HJ, Mumford E, Glass GV, et al. Mental health treatment and medical care utilization in a fee-for-service system: Outpatient mental health treatment following the onset of a chronic disease. Am J Public Health 1983; 73:422–429.

Renal Biopsy: Why and When

Lewis Reisman, m.d., Steven Dikman, m.d., Jacob Churg, m.d., and Sherman Kupfer, m.d.

Abstract

The purpose of this paper is to emphasize the importance of information obtained by renal biopsy in the diagnosis, prognosis, and therapy of patients with renal disease. Because controversy persists regarding the value of renal biopsy as an aid in determining prognosis and in choosing appropriate therapy, there has been some reluctance to use it early after the onset of obvious signs, symptoms, and laboratory findings indicative of renal disease with or without involvement of other organs. Although all such patients may not benefit from the information provided by a proper biopsy, we will illustrate some of the characteristic histologic details found in specific circumstances in our experience where the biopsy has been particularly helpful in reaching a diagnosis, in assessing prognosis, and in choosing therapy.

An enormous amount of information on the nature and prognosis of kidney diseases has been obtained from renal biopsy material. Despite this, controversy persists regarding the utility of renal biopsy not only with respect to diagnosis and prognosis but also in aiding in the choice of therapy. Consequently, there has been some reluctance to use it early during the initial workup. Simply stated, the reasons for performing a renal biopsy are: (a) providing a specific diagnosis; (b) defining severity and potential for reversibility of the histologic abnormalities; and (c) aiding in the choice(s) in management and therapy. Certainly, all patients with renal disease may not benefit from the information obtained from an adequate biopsy. The purpose of this paper is to ascertain when the information disclosed by a renal biopsy would be of greatest value in the care of a patient, adult or child, with renal disease and to illustrate some of the characteristic histologic details found in specific diseases.

The prelude to any consideration of renal biopsy must be the evaluation of a detailed medical history including recent illness, use or abuse of drugs, exposure to toxic substances at home or work, and family history. Finally, the clinical examination must be designed and executed to identify both local and systemic manifestations of renal disease and, thereby, to define the possible diagnosis. Although a whole spectrum of laboratory examinations and imaging studies will be suggested at this juncture, it is crucial to state that the urine must be examined directly by the physician. This examination of the urine includes a microscopic examination of a stained sample of the urinary sediment, either a smear or cytospin (using May-Grunwald Giemsa, Wright, or Diff-Quik staining procedures), because certain diagnostic morphologic features can not be appreciated in the ordinary unstained wet preparation.

Having identified the likely differential diagnostic possibilities, a renal biopsy should be considered in those patients with nephrotic syndrome as well as those who have moderate or nephrotic range proteinuria. In a subgroup of patients with nephrotic syndrome, the clinical diagnosis was wrong in 45% of the cases (1). Further, it is now known that some of these patients who have a form of focal segmental glomerulosclerosis might respond to long-term steroid therapy whereas others would be resistant. In addition, renal biopsy may be warranted in selected patients who manifest relatively recent and/or acute

---

From the Departments of Pediatrics (LR), and Pathology (SD, JC), and Medicine (SK), The Mount Sinai School of Medicine, New York, NY 10029. Address requests for reprints to Sherman Kupfer, M.D., Box 1100, The Mount Sinai School of Medicine, One East 100th Street, New York, NY 10029.

onset of renal insufficiency and concurrent involvement of other organs. It is of interest to note that the pathologic diagnosis was significantly different from the clinical diagnosis in more than 60% of the cases, whereas the course of treatment was altered in more than 30% (1). In a group of patients with acute onset of severe vascular involvement of glomeruli, the findings on renal biopsy would dictate the role and nature of urgent therapy such as immunosuppression and chemotherapy. Similarly, the extent of renal involvement in systemic lupus erythematosus cannot always be determined from serologic and clearance data, and/or from the urinalysis.

One might defer the use of steroids, immunosuppression, chemotherapy, or plasmapheresis, and avoid their attendant risks if the renal histology revealed IgA nephropathy, hypertensive nephrosclerosis, amyloidosis, congenital nephritides, some mild forms of membranous glomerulopathy, and glomerular thin basement nephropathy. On the other hand, the information provided by a renal biopsy is important before the formulation of a plan for the appropriate clinical management of patients who have had a renal transplant and who do not have a gross anatomical cause for deteriorating renal function.

The diagnostic value of the renal biopsy depends clearly on the adequacy and integrity of the sample. Guidance with ultrasound and/or computed tomographic (CT) scan, and the use of the 14-gauge needle ordinarily will provide an adequate sample. Hematuria, pain, and anemia were seen in less than 3% of those on whom a biopsy was performed when guidance was provided by ultrasound (2). Asymptomatic perirenal hematomas may be seen more commonly on postbiopsy ultrasound examination or with CT scan. Rarely, hypotension may be noted with more extensive bleeding. Another complication may be an arteriovenous fistula, which usually resolves spontaneously. A more recent innovation has been the use of 18-gauge automated needles and even smaller bore thin-walled needles. Although these needles are associated with fewer of the aforementioned complications, the adequacy of the single-core sample so obtained may be marginal. Multiple passes may overcome the sampling problem without necessarily increasing the frequency of complications. Irrespective of the size of the needle used, an experienced individual must examine the core immediately with a dissecting microscope to determine whether sufficient renal tissue has been obtained. The specimen must then be prepared for histologic, immunofluorescent, and electron microscopic techniques which are essential for an adequate examination.

As stated previously, the objective of this paper is to focus on the clinical indication for renal biopsy. Although we will cover several broad groups of renal diseases, the following examples are not intended to cover every differential diagnostic possibility and all sophisticated aspects of the pathological details that might be seen in any given biopsy.

Nephrotic Syndrome

Significant proteinuria, with or without the presence of nephrotic syndrome, is an accepted indication for kidney biopsy. Nephrotic syndrome is defined as the presence of heavy proteinuria, hypoalbuminemia, and edema. It is usually associated with hypercholesterolemia and may be associated with hematuria as well as other signs of renal disease such as decreased glomerular filtration rate. Possible causes of nephrotic syndrome and nephrotic range proteinuria include: a variety of drugs, infections, malignancies, and collagen-vascular disease, but the overwhelming majority of cases are idiopathic. The idiopathic disease entities have been assigned names that are descriptive of the pathologic features seen in the biopsy specimens. The natural history and prognosis of renal disease as well as the choice of therapy depends on the pathologic findings.

Minimal change disease (nil disease, lipoid nephrosis) is most commonly found in children with nephrotic syndrome. Membranous nephropathy is most common in adults. During the second decade of life membranous nephropathy increases and minimal change disease decreases in incidence and the former overtakes the latter (3).

In children between the ages of 2 and 6 years, nephrotic syndrome is relatively common and has an incidence of five per million. Of these cases, 85 to 90% will have minimal change disease on biopsy (4). These children will typically present with the sudden onset of edema. The edema may become quite massive within 1 to 2 days and include periorbital edema, ankle and leg edema, scrotal or labial edema, and massive ascites. The sudden and alarming appearance of massive edema will almost always cause the caregivers to bring the child for immediate medical attention. The affected child has typically been healthy previously, although he or she may have an atopic history with multiple allergies. The family history is noncontributory. Urinalysis will reveal massive proteinuria and little else. There may be a few erythrocytes in the sediment but gross or even significant microscopic hematuria would lead the clinician away from the diagnosis of minimal change nephrotic syndrome. A very low serum albumin level (usually less than 2.0 g/dL) as well as a low serum total protein is found. The serum creatinine level will be normal, although the blood urea nitrogen (BUN) may be somewhat elevated because of intravascular depletion.

Treatment of these patients begins with large amounts of diuretics, usually furosemide, assuming that intravascular depletion is not present. Sometimes the furosemide will not cause an adequate diuresis and zaroxolyn or aldactone will be added for their synergistic effect. Infrequently an infusion of 25% albumin will be given to mobilize the excess fluids from the interstitium so that they may be excreted.

Because minimal change is so ubiquitous in children with nephrotic syndrome between the ages of 2 and 6 years, we do not usually rush to perform a biopsy if they present with the typical features described above. The children will be evaluated for other causes of nephrotic syndrome, especially infection and malignancy and, if these are not found, the patient will be given an oral dose of prednisone, twice daily. The nephrotic syndrome will remit in almost all children with minimal change disease within 8 weeks after initiating this treatment (5). A lack of response to therapy would lead to the strong concern that the nephrosis is due to another disease with a worse prognosis. In this case, a biopsy would be indicated.

The essential morphologic feature of minimal change disease is effacement of podocyte foot processes by electron microscopy and the absence of other glomerular abnormalities (Fig. 1). The light microscopic examination with periodic acid-Schiff (PAS) stain is critical to examine for subtle glomerular and tubular scarring. Finding an even minute focus of tubular atrophy in a young patient should arouse suspicion of a more ominous scarring process, such as focal segmental glomerulosclerosis (FSGS; see below). A mild increase in mesangial matrix and a few mesangial deposits of IgM and/or C3 may be found in otherwise typical nil disease. It is crucial of course to obtain an adequate sampling of tissue because the absence of focal lesions in an inadequate biopsy does not exclude FSGS in patients who have had a poor response to therapy.

Although minimal change nephrotic syndrome goes into remission fairly readily with prednisone and the disease does not progress to renal failure, most patients will develop a pattern of relapses that may continue for many years. Complications of nephrotic syndrome, including thrombosis and a variety of infections, the most troubling of which is peritonitis, may develop. In addition, the chronic use of steroids may lead to a variety of secondary problems. Alkylating agents and cyclosporine have been used to spare prednisone and its side effects in those patients with very frequent relapses.

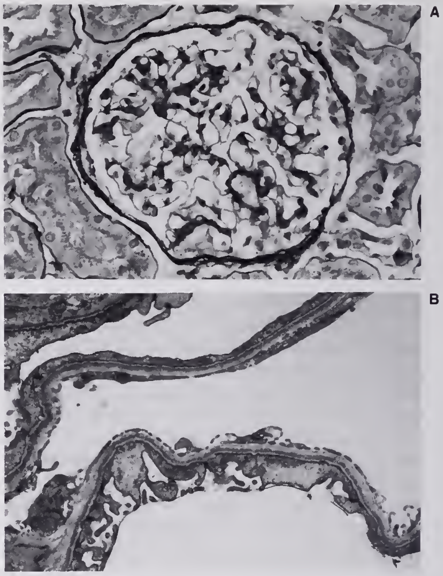


Fig. 1. Nil Disease (minimal change). (A) Photomicrograph of lipoid nephrosis illustrating a normal-appearing glomerulus by light microscopy (PAS). (B) Electron micrograph of glomerular capillary wall from a patient with minimal change nephrotic syndrome showing extensive effacement of podocyte foot processes.

Focal Segmental Glomerulosclerosis

In 15% of children with idiopathic nephrotic syndrome, the biopsy reveals FSGS. Approximately one third of these patients will respond to treatment with prednisone in a manner identical with those with minimal change. The prognosis of these patients is probably as benign as those with minimal change. The other two-thirds of patients with this diagnosis do not respond to a usual course of prednisone and their prognosis is considerably worse, with 50 to 100% of them fated to eventually have renal failure. Furthermore, in

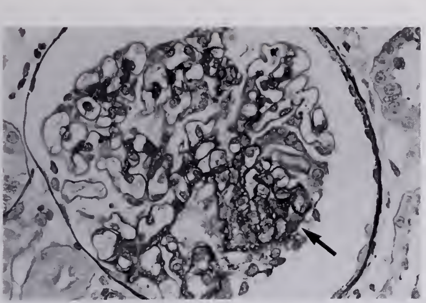


Fig. 2. Focal segmental glomerulosclerosis (FSGS). SM, a 12-year-old boy, was noted to have mild pretibial edema on a routine examination. The urinalysis revealed 3+ proteinuria and the sediment was unremarkable. The results of his physical examination were also unremarkable. Serum total protein and albumin levels were 4.6 g/dL and 2.0 g/dL, respectively. The serum cholesterol level was 409 mg/dL. This light microscopic photomicrograph of FSGS illustrates an otherwise unremarkable glomerulus with a segmental area (arrow) of capillary collapse and sclerosis with overlying enlarged podocytes (PAS).

20% of patients the disease will recur after transplantation (6).

Patients with this form of idiopathic nephropathy may present in a manner indistinguishable from minimal change disease, with their disease only being identified when they do not respond to prednisone. However, these children often are older than those who have typical minimal change nephrotic syndrome. Moreover, nephrotic syndrome may have begun gradually and insidiously rather than suddenly. In some cases there may be no edema at all, and only the incidental finding or proteinuria or some hypoalbuminemia may be evident. The prognosis of such patients is undetermined. Some will remain quite well, whereas others will later develop nephrotic syndrome.

The urinalysis of patients with FSGS usually shows only proteinuria with very little sediment. Hypertension, when present, will indicate more advanced disease, as will an abnormally high serum creatinine level.

Those patients with FSGS who did not respond to usual doses of prednisone, in the past, were treated with diuretics alone. More recently other treatments have been tried and are beginning to gain support among nephrologists. Mendoza et al. (7) recommended that unresponsive patients be treated with intermittent intravenous doses of methylprednisolone as large as 30 mg/kg and cyclophosphamide. Others have recommended treatment with cyclosporine (8). This treatment remains controversial even though cyclosporine has been clearly demonstrated to decrease or eliminate proteinuria in these patients, at least temporarily, and may lead to a long-term benefit. There is speculation that hypercholesterolemia may play a role in the progressive damage of the kidneys caused by this disease and that drugs and diet that decrease the serum cholesterol may be of benefit (9). Finally, renal failure may be delayed by decreasing hyperperfusion and hyperfiltration in the damaged kidneys by reduction in dietary protein intake and by administration of an angiotensin-converting enzyme inhibitor.

The light microscopy, immunofluorescence, and electron microscopic morphologic findings are generally similar to minimal change nephropathy in the early or milder forms of the disease, except that foot process effacement tends to be less diffuse. The characteristic appearance by light microscopy is a focal (some glomeruli) and segmental (parts of glomeruli) scarring process without proliferation of cells that characterizes glomerulonephritis (Fig. 2). The segmental glomerular basement membrane collapse and sclerosis is usually accompanied by overlying hypertrophied podocytes, which may appear as proliferation in sections stained with hematoxylin and eosin (H&E). The scarring tends to be more common in glomeruli in the deeper cortex so that superficial biopsies may miss these lesions. Even with small or superficial biopsies, a tell-tale sign that the nephrosis is not nil disease is the presence of minute foci of tubular atrophy caused by nephrons that have scarred because of the progressive glomerular changes. The sclerosing glomerular changes are often accompanied by IgM and C3 deposits that appear to represent trapping of serum proteins rather than an immunologic process.

At present, AIDS nephropathy should be suspected when FSGS is accompanied by extensive glomerular collapse, tubular microcystic changes, marked interstitial inflammation, and tubuloreticular structures in endothelial cells and/or lymphocytes.

The causes of minimal change and focal and segmental glomerulosclerosis and their relationship remain unknown. Many patients have undergone renal biopsies twice, once at the onset of their nephrotic syndrome and a second time when their pattern of response to prednisone had changed. Sometimes, the morphologic features have changed in the intervening period, nil disease being diagnosed in the first biospy and FSGS in the second biopsy. It is unknown whether minimal change can evolve in FSGS or whether, because of the focal nature of the lesion, FSGS was simply missed in the first biopsy.

Membranous Glomerulonephropathy

Membranous nephropathy presents with significant but otherwise asymptomatic proteinuria, even nephrotic syndrome, and some microscopic hematuria in patients (male preponderance) usually older than 30 years. The finding of membranous nephropathy on biopsy usually leads to extensive evaluation in the search for underlying neoplasia, diabetes mellitus, hepatitis B and other less common infectious and parasitic diseases, systemic lupus erythematosus (SLE), and exposure to a variety of drugs or chemicals. Nonetheless, the cause in most of these patients is never identified and is termed idiopathic.

In contrast to minimal change disease and FSGS, the glomerular capillary walls are thick and rigid with round glomerular lumens (Fig. 3). This reflects the presence of subepithelial deposits, almost invariably accompanied by varying degrees of scarring on the epithelial aspect of the glomerular basement membrane (GBM). The scarring appears early as spike-like projections of basement membrane material associated with the subepithelial deposits in all but the rarely encountered earliest forms of the disease. The glomerular capillary wall scarring is usually slowly progressive correlating with the generally slow clinical progression to renal failure. Mesangial deposits are not characteristic of idiopathic membranous nephropathy. The presence of atypical features such as deposits in the mesangium or at other sites in the glomerulus, as seen by immunofluorescence and/or electron microscopy, should lead one to suspect underlying SLE or a variety of other causes listed previously, especially viral hepatitis.

Membranoproliferative (Mesangiocapillary) Glomerulonephritis

Patients with membranoproliferative glomerulonephritis (MPGN) (sometimes referred to as mesangiocapillary glomerulonephritis) tend to be younger than those with membranous glomerulonephritis. They usually present with significant proteinuria and nephrotic syndrome, recurrent gross or microscopic hematuria, evidence of decreased renal function, hypertension, and in many cases, hypocomplementemia, particularly when neither clinical nor serologic findings of SLE or recent bacterial/viral infection are evident. Of particular interest has been the finding of cryoglobulinemia in association more commonly with hepatitis C than with hepatitis B. Establishing the diagnosis with a renal biopsy is important in assessing the likely response to appropriate therapy.

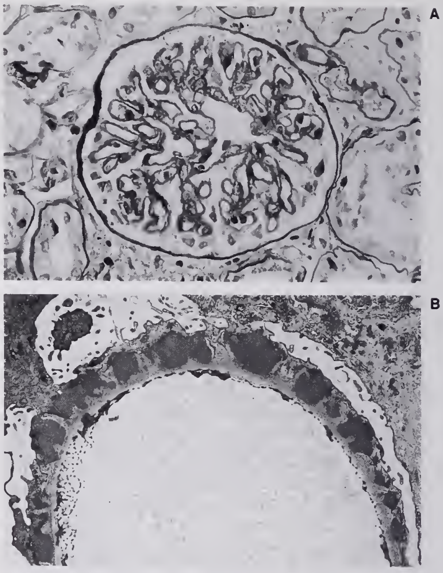


Fig. 3. Membranous nephropathy. The patient is a 49-year-old man with a history of nephrotic syndrome spanning several years. For the past 9 months he had been experiencing shortness of breath, associated with a left pleural effusion. There had been some increase in edema during the past 6 months. Serum creatinine and BUN levels were 2.7 and 48 mg/dL, respectively. The ANA level, hepatitis B and hepatitis C markers were all negative. Serum albumin level was low and the cholesterol level was increased. (A) Photomicrograph of glomerulus showing diffuse glomerular basement membrane thickening and normal mesangial regions (PAS). (B) Electron micrograph of a glomerular capillary wall with numerous subepithelial electron dense deposits, some associated with thin lamella of basement membrane-like material forming "spikes."

The glomerular changes in MPGN involve both proliferation of mesangial cells, which indicates an active component, and a chronic organizing component represented by increased mesangial matrix (Fig. 4). The proliferative and sclerosing process expands beyond the usual confines of the mesangium into the adjacent capillary walls. The route of extension is between the glomerular basement membrane and endothelium forming the characteristic complexity and double contours of the capillary walls often accompanied by subendothelial deposits. The combined proliferative and sclerosing glomerular response to injury corresponds to the usual chronic, progressive, and unresponsive clinical course. MPGN is a pattern of glomerular response to a wide range of diseases including SLE, cryoglobulinemia, hepatitis C, and IgA nephropathy, with

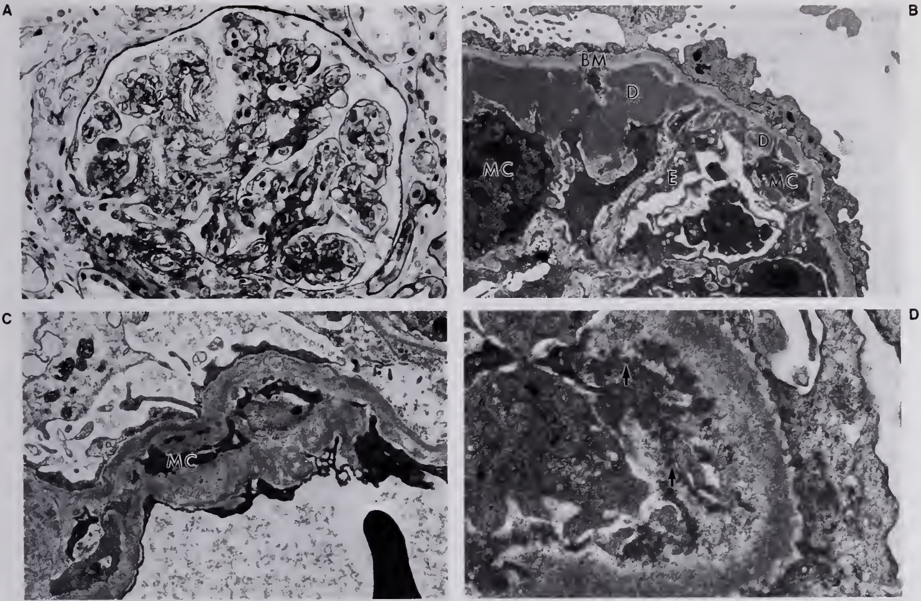


Fig. 4. Membranoproliferative (mesangiocapillary) glomerulopathy. RM is a 44-year-old man with cirrhosis thought to be caused by hepatitis C. Two years ago he had been treated with interferon. During the ensuing 2 years, however, liver function has deteriorated. Five months before this admission he was noted to have proteinuria (5 g/day) and blood pressure elevated to 160/90 mm Hg. There were ascites. Serum albumin was 2.4 g/dL, while serum creatinine and BUN levels were 1.2 and 16 mg/dL, respectively. Because the patient was a candidate for a liver transplant, concern was expressed regarding progression of renal disease caused by cryoglobulin deposition within the kidney and whether a simultaneous renal transplantation should be performed. (A) Photomicrograph of glomerulus with accentuated lobules containing multiple capillary walls with double outlines, mesangial proliferation, and increased matrix (PAS). (B) Electron micrograph of glomerular capillary wall from a patient with SLE showing large subendothelial deposits (D) and extension of mesangial cell (MC) between the endothelium (E) and basement membrane (BM) narrowing the capillary lumen. (C) Electron micrograph from a patient with mesangiocapillary glomerulonephritis showing extension of mesangial cells (MC) into a glomerular capillary wall. (D) Electron micrograph of cryoglobulin deposits in the subendothelial aspect of the glomerular basement membrane. The deposits are composed of tubular structures cut longitudinally and in cross-section (arrow).

each disease process having additional clinical and morphologic features that facilitate its identification. The typical ultrastructure of cryoglobulin deposits consists of cylindrical structures seen longitudinally and in cross-section. Their presence should alert one to the diagnosis of cryoglobulinemic glomerulonephritis.

Acute Renal Insufficiency

Rapidly Progressive Glomerulonephritis (Crescenteric GN)

A biopsy does not need to be performed on all patients who have what appears to be acute glomerulonephritis from one or another cause. However, those generally normotensive anemic patients with some renal insufficiency, who experience steady or even sometimes alarmingly rapid deterioration which progresses to renal failure, must have a renal biopsy if otherwise not contraindicated. The timely institution of appropriate therapy may reverse the process, and/or restore and preserve an adequate level of renal function. In this group of patients, for example, are those with anti-GBM antibody disease (Goodpasture's syndrome and Goodpasture's disease), pauci-immune glomerulonephritis (ANCA positive, Wegener's granulomatosis), and other forms of vasculitis.

In these forms of rapidly progressive glomerulonephritis (RPGN), the glomerular tuft is quite normal except for the presence of a few small breaks in the GBM. Fibrin and possibly other substances will leak into Bowman's space through these breaks and elicit a florid, almost explosive, cellular response. Numerous mononuclear cells migrate and proliferate in Bowman's space forming a mass of cells shaped like a crescent moon. In the early and possibly reversible stage, the crescents are cellular but quickly change to fibrocellular and then to fibrous scarring (Fig. 5). The proliferation and scarring in Bowman's space rapidly compress and destroy the glomerulus and attached nephron. The renal biopsy helps to quickly identify this type of explosive and devastating glomerular response to injury. Prompt initiation of appropriate treatment is essential if renal function is to be preserved.

Crescents also may be found in many other forms of severe acute glomerular injury such as the immune complex diseases, but the clinical course of these diseases, even when accompanied by small crescents, is often more indolent than the vasculitides and ANCA-related entities noted above.

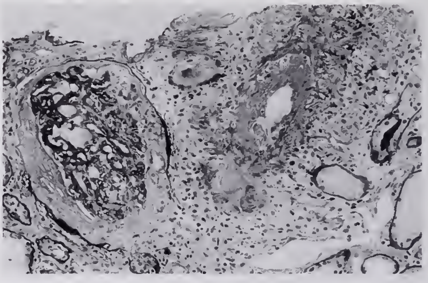


Fig. 5. Rapidly progressive glomerulonephritis (Wegener's granulomatosis). The patient is a 76-year-old woman with a long history of rheumatoid arthritis. She also had severe episodes of episcleritis which required maintenance steroid therapy as well as nonsteroidal medication. The patient developed a large gastric ulcer that responded to H2 blockers from which she developed thrombocytopenia. More recently, she was found to have evidence of interstitial lung disease. She manifested increasing weakness, lost 50 pounds, and had a new onset of microscopic hematuria. The serum creatinine level was 1.8 mg/dL. About 6 weeks later, on admission to the hospital, the urinary sediment revealed dysmorphic red blood cells, red blood cell casts, and granular casts. The BUN and the serum creatinine levels were 87 and 9 mg/dL, respectively, whereas the serum potassium was 6.9 mEq/L. Hemodialysis was started. The results of a bleeding and coagulation workup were normal. A renal biopsy was performed, after which she developed progressive infiltrates in both lung fields. These infiltrates were accompanied by bouts of hemoptysis. The serum ANA and ANCA levels were reported thereafter as being (antimyeloperoxidase, "low positive"). The patient died 1 week after the renal biopsy. This photomicrograph shows glomerulus (left) with a large fibrocellular crescent filling Bowman's space and associated with collapse of the glomerular tuft. On the right is a small artery with granulomatous vasculitis (PAS).

Hemolytic Uremic Syndrome

Hemolytic uremic syndrome (HUS) is diagnosed by the combination of microangiopathic hemolytic anemia, thrombocytopenia, and renal failure with anuria or oliguria. The condition generally presents in epidemic form in toddler-age children during the spring or summer seasons after severe gastroenteritis and diarrhea, often bloody, which may lead to dehydration. The disease has been associated with a specific pathogenic strain of *Escherichia coli* (10), but other pathogens have been incriminated as well. The affected child will have bloody urine, but may quickly develop oliguria or even anuria and renal failure. Hypertension is common and seizures may occur either from hypertensive encephalopathy or from the disease process itself affecting the central nervous system. These findings together with the microangiopathic hemolytic anemia and thrombocytopenia confirm the diagnosis. A biopsy is rarely indicated (and often contraindicated) early in the disease. A biopsy ultimately may be indicated to determine prognosis if complete recovery of renal function is delayed or does not occur.

In addition to the typical presentation in toddlers described above, this disease may present in patients of any age without accompanying gastroenteritis. The disease may present in adults secondary to an infection, as a result of a complication of pregnancy, or during the postpartum period, and as a complication of contraceptive agents. There also is a familial form of HUS which is not seasonal and is not associated with gastroenteritis. This familial form of HUS has a more indolent, recurring course and an ultimately worse prognosis. The very similar disease of thrombotic thrombocytopenic purpura may be a variant of HUS.

A microangiopathic glomerular response characterized by an alteration in the glomerular capillary endothelium is common to a group of diseases including HUS. The renal biopsy helps to identify these uncommon diseases with often complex multisystem manifestations and provides information as to the severity and activity of the process. In contrast to glomerulonephritis, there is little or no proliferation of the mesangium and immunologic type deposits are not generally found by immunohistologic or electron microscopic techniques. The subendothelial aspect of the glomerular capillary walls and mesangium are widened by pale reticulated material (Fig. 6). Fibrinogen may be a component of these fluffy, reticulated regions but only edema and nonim-

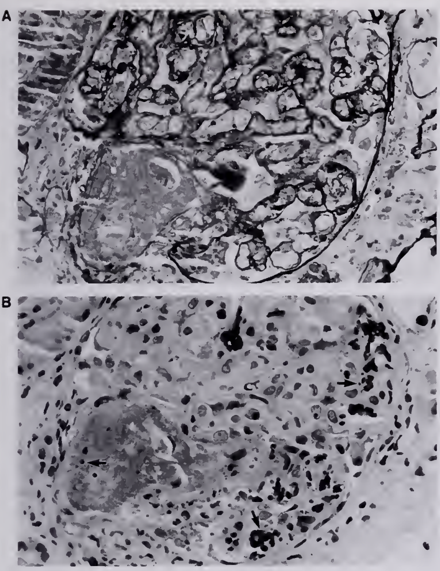


Fig. 6. Hemolytic uremic syndrome. (A) Photomicrograph of glomerulus with a dilated and aneurysmal afferent arteriole containing fibrin. Some of the glomerular capillary walls have an irregular reticulated endothelial aspect and flocculent mesangial regions (PAS). (B) The same glomerulus showing hypocellular pale appearance with red blood cell fragments (arrows) in the capillary walls and dilated afferent arteriole (H&E).

munologic proteinaceous material usually are identified. Erythrocytes become trapped and torn in the lumen of deformed glomerular capillaries during their passage through the microcirculation. These fragmented erythrocytes accumulate locally and then can be found circulating in the blood. If the microangiopathic process is severe and if larger vessels are also involved in this process, varying degrees of acute renal failure may become apparent clinically. The microangiopathic process generally leads to scarring and some permanent renal damage. Its severity depends on the diffuseness of the glomerular involvement and size of the involved vessels.

There has been considerable progress in our understanding of the pathophysiology of HUS. Disseminated intravascular coagulation clearly plays a major role. A release of endotoxin (i.e., from bacteria in the gastrointestinal tract) may result in vascular endothelial damage, deposition of fibrin, destruction of erythrocytes, damage to platelets, and the release of thromboplastin. The increased levels of thromboplastin would cause further deposition of fibrin and continued disease.

Over the years many therapies have been suggested to shorten the course and improve the outcome of HUS. Each, in turn, has been proven ineffective. The only success has been with close monitoring and aggressive treatment to maintain fluid and electrolyte balance, blood pressure control, and an adequate hematocrit. Dialysis, when required, should be initiated early.

Tubulointerstitial Nephritis

Tubulointerstitial nephritis is not an infrequent adverse reaction to many drugs in common use such as penicillin, methicillin, cephalosporins, nonsteroidal anti-inflammatory drugs, and rifampicin. It has been recognized as an important cause for acute renal insufficiency, often presenting with oliguria, hematuria, proteinuria, and even the nephrotic syndrome. Thus, it is often difficult to distinguish the clinical presentation of tubulointerstitial nephritis from that of rapidly progressive glomerulonephritis, acute renal failure due to an untoward hemodynamic event, sepsis, or direct drug nephrotoxicity. The presence of red cell casts and dysmorphic erythrocytes in the urinary sediment generally rules out tubulointerstitial disease, but the absence of these elements on repeated phase-contrast microscopy of the urinary sediment does not rule out glomerular disease.

Renal interstitial inflammation is often a component of primary glomerular and vascular disease especially when the glomerular and vascular changes are severe and advanced. However, tubulointerstitial nephritis is characterized by an inflammatory process of the renal interstitium out of proportion to glomerular or blood vessel changes. The cellular infiltrate can range from acute to chronic type inflammatory cells. Although the presence of eosinophils is characteristically found in allergic type conditions, it nonetheless can be a nonspecific finding in many forms of tubulointerstitial nephritis. The inflammatory cells often permeate tubular walls (tubulitis) which may result in destruction of the tubular basement membranes and irreversible scarring. Somewhat similar histologic findings are found in cellular rejection (see Fig. 9).

Acute Glomerulonephritis (Postinfectious)

The most common cause of hematuria in children is urinary tract infection but the second most common cause is postinfectious acute glomerulonephritis (AGN). Classically, 2 to 3 weeks after an infection with a nephritogenic strain of streptococcus, the patient, usually a school-age child between 6 and 9 years of age, will present with edema and hematuria. In northern climates the streptococcal infection will more often be in the form of a tonsillitis. In warmer areas of the country the infection will usually be in the form of impetigo. Cola-colored, reddish brown, or smoky tea-colored urine are found in 70% of affected patients, all the rest presenting with microscopic hematuria. The edema may be mild or severe but is usually not due to nephrotic syndrome and hypoalbuminemia. Hypertension may be present and occasionally can be severe enough to cause encephalopathy and heart failure. There is usually some degree of renal insufficiency as evidenced by increased serum levels of both creatinine and urea nitrogen. The serum sodium level may be low and the serum potassium high. Rarely, renal failure will be severe enough to require dialysis. The major risks to the life of these patients are hypertension and hyperkalemia. The C3 component of complement is low, albeit transiently, in 90 to 100% of patients. Serum titers of antistreptolysin and other antibodies to streptococcal antigens usually are elevated. The renal abnormalities and the serum complement will return to normal within 6 to 8 weeks (11), although the patient may have persistent microscopic hematuria for 6 to 12 months and indeed may have recurrent gross hematuria with intercurrent infections. A renal biopsy usually is not indicated to make the diagnosis. Bad prognostic signs and indications for biopsy would include persistent disease and low complement for more than 8 weeks, or persistent proteinuria and hypertension after resolution of the acute disease. When present, streptococcal infection should be treated. In addition, therapy should be provided to control and manage hypertension, hyperkalemia, and edema should these manifestations occur. There is no treatment for the disease process itself.

Postinfectious glomerulonephritis is characterized by proliferation of mononuclear cells (mesangial cells and macrophages) within the glomerular tuft as an acute, usually completely reversible reaction to an immunologic agent. The edema and cellular proliferation within the glomerular tuft compress the adjacent glomerular capillaries (Fig. 7) and may be responsible for the oliguria and hypertension seen clinically. Minute holes in the basement membrane have been found by electron microscopy. These holes allow passage of red blood cells into the tubules (red blood cell casts) and urine (hematuria). These GBM defects do not appear to cause permanent injury to the jelly-like glomerular basement membranes. The glomerular changes in AGN are not accompanied by an increase in mesangial matrix or scarring of the capillary walls characteristic of most other forms of glomerulonephritis. When the initiating immunologic agent is gone in AGN, the extensive diffuse acute proliferative changes generally resolve completely with reopening of the compressed capillaries and return of renal function. A renal biopsy may assist in the management of a patient with atypical clinical features in an otherwise typical acute postinfectious nephritic syndrome. The finding of typical AGN morphologic changes may obviate the need for the use of potentially toxic therapeutic modalities. In more severe and atypical clinical forms of an acute nephritic syndrome, a renal biopsy may aid in assessing the presence of crescents, necrosis, thrombosis, larger vessel involvement and other changes which may warrant a modification of therapy. Acute proliferative glomerulonephritis with considerable focal and segmental variation rather than the typically diffuse appearance of postinfectious glomerulonephritis should prompt consideration of a variety of systemic diseases such as viral infections and SLE. The renal biopsy in an acute nephritic patient may also aid in identifying early forms of MPGN and nonimmunologic glomerular injury (e.g., HUS) which have different prognoses and treatments.

It is of interest to note that recently obtained biopsy material illustrating AGN after streptococcal infection is not readily available. It is important to point out that typical and atypical forms of acute proliferative glomerulonephritis have been encountered increasingly with viral infections, particularly hepatitis C. Hepatitis C, often accompanied by cryoglobulinemia, has been associated with a variety of clinical and morphologic manifestations of glomerulonephritis including AGN and MPGN often with characteristic ultrastructural deposits shown previously.

IgA Nephropathy

IgA nephropathy or Berger's disease is a much more common disease than realized originally. The etiology of this condition remains unknown. Its presentation is variable and the disease may present with varying amounts of hematuria and/or proteinuria. The prognosis of this disease is also variable with the adult form often progressing to renal failure. Whether the childhood form is more benign is not entirely clear because many of the patients continue to be followed as adults by their internist or nephrologist.

Typically, the childhood form of the disease presents with episodes of gross hematuria, often

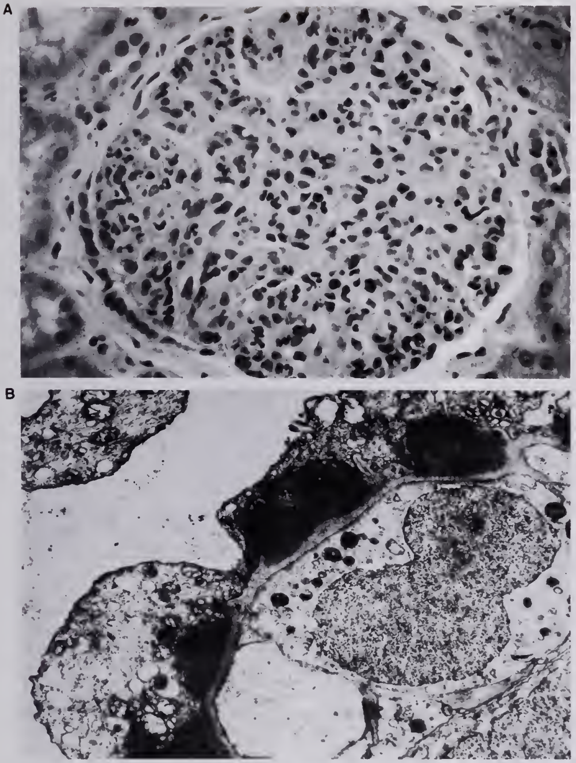


Fig. 7. Acute post infectious glomerulonephritis. (A) Photomicrograph showing marked hypercellularity involving the entire glomerulus with obscured capillary lumens caused by edema and hypercellularity (H&E). (B) Electron micrograph of a glomerular capillary wall with large subepithelial deposits associated with effacement of the overlying podocyte foot processes. A leukocyte is in direct contact with the GBM.

associated with intercurrent infections lasting from one to several days. Gross hematuria may be associated with proteinuria. Persistent proteinuria and microscopic hematuria in between episodes of gross hematuria is associated with more malignant disease and a worse prognosis. The nephrotic syndrome is rare in the childhood form of IgA nephropathy. Persistent microscopic hematuria in children is frequently caused by a mild form of IgA nephropathy (12).

In adults, IgA nephropathy will often present as a mixture of nephrotic syndrome with edema and hypoalbuminemia and nephritis with hematuria and hypertension and a clinical course leading to end-stage kidney disease. The use of prednisone and cyclophosphamide in the treatment of IgA nephropathy remains controversial with the weight of opinion being that it is not useful. There is some evidence that fish oil may be beneficial (13).

The glomerular involvement in Berger's disease characteristically shows considerable variation from glomerulus to glomerulus (Fig. 8). The usual appearance is that of a chronic mesangial or mesangiocapillary glomerulonephritis pattern with increased mesangial matrix accompanied by IgA deposits seen in immunopathologic studies. The deposits may often be quite large and visible by light microscopy as hyaline material in the mesangium and/or glomerular capillary walls. The activity may range from acute to chronic forms of glomerulonephritis.

Henoch-Schönlein purpura (HSP) has many features that are difficult to distinguish from IgA nephropathy. Indeed, some have suggested that IgA nephropathy is a localized variant of HSP. This condition presents in children and is diagnosed by a typical purpuric rash on the lower extremities, buttocks, and forearms. Most patients also suffer from arthritis in one or more large joints and bloody colic. Any organ system may be involved. Approximately 40% of patients will have renal involvement with hematuria and proteinuria. The prognosis of these patients will depend on the course of the renal disease. HSP is generally characterized by focal glomerular necrosis which is not a common feature of IgA (Berger's) disease. Although most patients with HSP will recover completely, approximately 5% will have chronic proteinuria and persistent hypertension. Many of these patients will progress to renal failure.

Post Renal Transplantation

There are two groups of patients in the immediate post-transplant period who are candidates for renal biopsy. In the first group are those patients who suddenly or gradually develop signs of renal insufficiency after a variable period of initially good renal function. Included in this diverse group of patients are those with delayed hyperacute vascular rejection, accelerated acute vascular rejection, acute cellular rejection, and cyclosporine and tacrolimus (FK506) toxicity. In general, and in the absence of elevated plasma levels of cyclosporine or tacrolimus, the therapeutic approach has been to initiate more aggressive immunosuppression, because it is more likely that this decrease in renal function is caused by rejection, particularly when either an anatomic or surgical cause and sepsis is not evident. If examination of the urine sediment does not confirm the presence of cellular rejection and fine needle aspiration biopsy is not available, a core renal biopsy is indicated to establish the presence of rejection if further aggressive immunotherapy is to be contemplated. In the second group are those patients whose transplanted kidney function has been persistently poor. The clinical impression is generally that of acute tubular necrosis in the absence of any other anatomic or surgical cause. These patients require maintenance hemodialysis during this interval of inadequate renal function. Despite this prolonged delay in renal recovery, susceptibility to possible superimposed acute cellular rejection is not diminished. In this situation, clinical signs of such rejection may not be readily apparent particularly when monoclonal (OKT3) or polyclonal antibody (ATG or ALG) has been administered as part of induction therapy. Nonetheless, aggressive immunosuppression may be indicated. Such a patient in whom both the urine sediment and biopsy revealed the onset of rejection is illustrated in Fig. 9.

Clinical evaluation alone does not yield much insight into the cause of gradual and/or progressive loss of renal function starting at some period after reasonably good initial post-transplant renal function. In these patients, only a renal biopsy may resolve the question of reactivation of the original primary disease, or even another disease involving the transplanted kidney, or chronic rejection. The choice of treatment obviously depends the histologic findings.

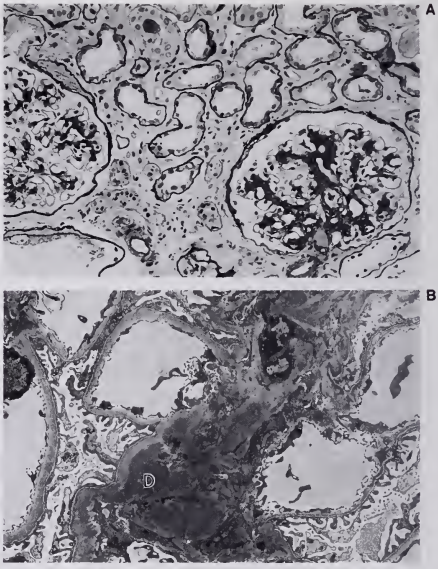


Fig. 8. IgA nephropathy. BR is a 57-year-old man who noted the onset of edema several months ago. Examination revealed cardiomyopathy. The urinary sediment was reported to be active. Serum ANA, cryoglobulin, complement, and all markers for hepatitis were negative. The serum creatinine level was 3.5 mg/dL. (A) Photomicrograph of two glomeruli with increased mesangial matrix that varies between and within the glomeruli. Tubular atrophy and interstitial fibrosis indicates a chronic process (PAS). (B) Electron micrograph of a portion of a glomerulus containing mesangial deposits (D).

Alport's Syndrome

Familial nephritis is a hereditary X-chromosome-linked disease presenting classically in affected males in the first decade of life with hematuria with or without proteinuria and bilateral sensorineural deafness for high frequencies. Recurrent gross or persistent microscopic hematuria is the most common presenting feature. Some proteinuria also may be found although rarely will it be massive and/or accompanied by the nephrotic syndrome. By the middle of the second decade of life these boys usually have renal failure requir-

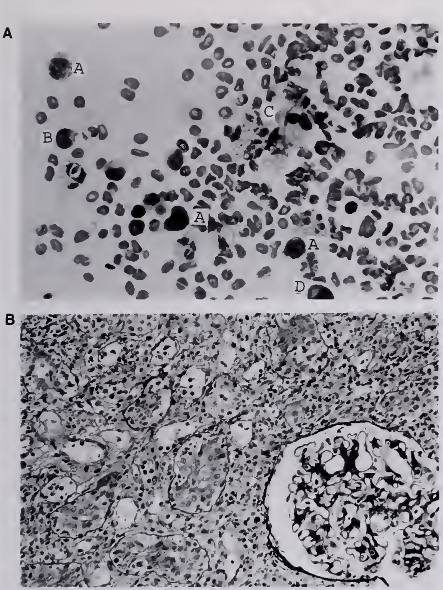


Fig. 9. Post-transplant rejection. AF is a 33-year-old man with a longstanding history of obstructive uropathy and repeated urinary tract infections. He had been on maintenance hemodialysis for 28 months. Moderately severe hypertension was being treated with clonidine and nifedipine. A cadaveric kidney transplant was performed after which he received monoclonal antibody immunosuppression for 10 days. The clinical impression was that of acute tubular necrosis. Renal function did not improve, and urine output remained low. Hemodialysis was required. Three weeks after transplantation, and after cyclosporine therapy had been started several days earlier, cytologic examination of the urine revealed lymphoblasts, lymphocytes, eosinophiles, and desquamated degenerating and vacuolated renal tubule cells and casts, all suggestive of a florid acute rejection. A core biopsy was obtained. (A) Urine sediment cytology with Diff-Quik stain shows lymphoblasts (A), lymphocytes (B), desquamated vacuolated renal tubule cells (C), and plasma cell (D). Not shown is the presence of eosinophils. (B) Photomicrograph shows marked interstitial and tubular infiltration by lymphoid cells with a normal-appearing glomerulus (PAS).

ing dialysis and are quite deaf. Their heterozygous mothers and affected sisters of the patient will typically have only microscopic hematuria and if they do progress to renal failure it will only be after their childbearing years. There is considerable variability to the classic pattern described above, and in some families the women and men are equally affected. In other families nerve deafness may be absent and there may be associated ocular findings. There are families in which the disease is clearly inherited in an autosomal recessive rather than X-linked pattern. Many sporadic cases have been seen with some or all of the features of Alport's syndrome but with no family history whatsoever.

Until recently Alport's syndrome could only be diagnosed by finding typical glomerular basement membrane changes on electron microscopic examination of the renal biopsy (Fig. 10). More recently, the cause of Alport's syndrome has been identified as a mutation in a gene on the X-chromosome responsible for the synthesis of type IV collagen in the glomerular basement membrane (14). An antibody now has been produced that binds to an epitope found in normal glomerular basement membranes. This epitope is absent in the glomerular basement membrane of patients with Alport's syndrome. There is no known treatment to prevent or delay the onset of renal fail-

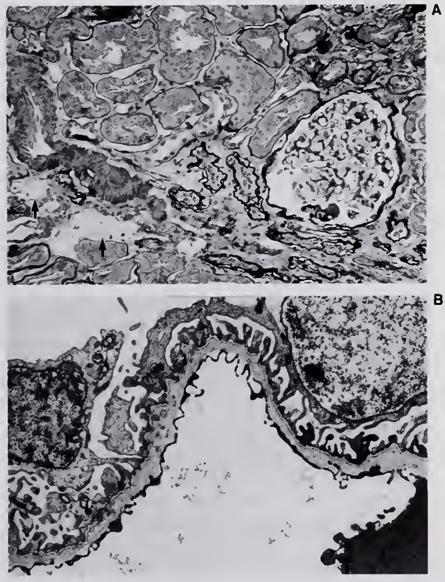


Fig. 10. Alport's syndrome. RN is a 47-year-old man with a long history of proteinuria (2.2. g/day), hematuria (3 +), and hypertension for which he was receiving Catapres patch and Procardia. The patient works in a chemical plant and is exposed to many toxins. Serum creatinine and BUN levels were unremarkable. Creatinine clearance was measured as 78 mL/min. A few small renal cysts were demonstrated on a renal sonogram which was otherwise unremarkable. His brother was receiving dialysis for end-stage renal disease of unknown type, possibly hypertension. (A) Photomicrograph of a glomerulus shows no significant changes by light microscopy. There is focal tubular atrophy and a cluster of interstitial foam cells (arrows) (PAS). (B) Electron micrograph of glomerular capillary wall shows marked remodeling by thin lamella of basement membrane material associated with scattered small dense granules. There are small areas of podocyte foot process effacement.

ure. These patients do not have an increased risk for dysfunction of a renal transplant.

Often it is clinically difficult to distinguish Alport's syndrome and benign familial hematuria. Benign familial hematuria may present with either microscopic hematuria or intermittent gross hematuria but without significant proteinuria. This condition is usually inherited as an autosomal dominant trait (15). There is neither associated hearing loss nor eye changes, and these patients do not develop renal failure. The finding of a close elderly relative of the affected patient who also has hematuria of long standing but no other symptoms or signs of renal disease is a very good prognostic sign.

Hematuria

Persistent and/or recurrent episodes of hematuria in both children and adults generally is brought to the attention of a nephrologist. Rarely, however, is hematuria alone an indication for a kidney biopsy. The patient should be evaluated by obtaining a history, physical examination, urine culture, careful microscopic examination of the urine, and urine and serum chemistries and followed by imaging examinations. These tests will usually suffice to determine whether the blood is coming from the kidneys or other parts of the urinary tract. Even if the blood is determined to be coming from the kidneys and is not the result of tumor or trauma, a kidney biopsy rarely will contribute to either prognosis or therapy. If hematuria is accompanied by significant proteinuria, recurrent gross hematuria, recurrent bouts of loin pain, bilateral nerve deafness, hypertension, low serum complement, positive antinuclear antibody (ANA) levels, or a significant family history, a biopsy will be necessary to make a definitive diagnosis. A wide range of specific morphologic diagnoses have been found in renal biopsies from this group of patients. Some patients with otherwise typical thin basement membrane nephropathy in which the GBM is less than 200 to 250 nM in thickness develop hypertension and renal insufficiency and thus should not be considered entirely "benign" (16). In a recent but uncontrolled study (17), the use of angiotensin-converting enzyme inhibitors even in patients without hypertension has reduced the frequency of recurrent episodes of loin pain and hematuria.

Acknowledgment

We are indebted to Ms. Sharon Smith for her superb photographic technical assistance.

References

1. Cohen AH, Mast CC, Adler SG, Kopple JD. Clinical utility of kidney biopsies in the diagnosis and management of renal disease. Am J Nephrol 1989; 9:309–315.
2. Meola M, Barsotti G, Cupisti A, Buoncristiani E, Giovannetti S. Free-hand ultrasound-guided renal biopsy: Report of 650 consecutive cases. Nephron 1994; 67:425–430.
3. Malick N, Short CD, Manos J. Clinical membranous nephropathy. Nephron 1983; 34:209–219.
4. Wyatt RJ, Marx MB, Kuzee M, Holland NH. Current estimates of the incidence of steroid responsive idiopathic nephrosis in Kentucky children, 1–9 years of age. Int J Pediatr Nephrol 1982; 3:63–65.
5. International Study of Kidney Disease in Children. The primary nephrotic syndrome in children. Identification of patients with minimal change nephrotic syndrome from initial response to prednisone. J Pediatr 1981; 98:561–564.
6. Leumann EP, Briner J. Recurrence of the primary disease in the transplanted kidney. In: Fine RN, Gruskin AV, eds. End state renal disease in children. Chap. 37. Philadelphia: WB Saunders, 1984: 528.
7. Mendoza SA, Reznick VM, Griswald WR, Krensky AM, Yorgin PD, Tune BM. Treatment of steroid resistant focal and segmented glomerulosclerosis with pulse methylprednisolone and alkylating agents. Pediatr Nephrol 1990; 4:303–307.
8. Lieberman KV, Tejani A. A randomized double-blind placebo controlled trial of cyclosporine in steroid-resistant idiopathic focal segmental glomerulosclerosis in children. J Am Soc Nephrol 1996; 7:56–63.
9. Grone EF, Walli AK, Grone HJ, Miller B, Seidel D. The role of lipids in nephrosclerosis and glomerulosclerosis. Atherosclerosis 1994; 107:1–13.
10. Spika JS, Parsons JE, Nordenberg D, Wells JG, Gunn RA, Blake PA. Hemolytic uremic syndrome and diarrhea associated with Escherichia coli 0157:#7 in a day care center. J Pediatr 1986; 109:287–291.
11. Ortiz JS, Finkle JF, Potter EV, Poon-King T, Ali D, Earle DP. Endemic nephritis and streptococcal infections in South Trinidad: surveillance studies during the first year following a major epidemic. Arch Intern Med 1970; 126:640–646.
12. Trachtman H, Weiss RA, Bennett B, Greifer I. Isolated hematuria in children: Indications for a renal biopsy. Kidney Int 1984; 25:94–99.
13. Hagg RJ. A randomized, placebo-controlled, multicenter trial evaluating alternate-day prednisone and fish oil supplements in young patients with immunoglobulin A nephropathy. Am J Kidney Dis 1995; 26:792–796.
14. Barker DF, Hostikka SL, Zhou J, et al. Identification of mutations in the COL4 A5 collagen gene in Alport syndrome. Science 1990; 248:1224–1226.
15. Eisenstein B, Stark H, Goodman RM. Benign familial hematuria in children from the Jewish communities of Israel: Clinical and genetic studies. J Med Genet 1979; 16:369–372.
16. Nieuwhof C, Doorenbos C, Grave W, et al. A prospective study of the natural history of idiopathic nonproteinuric hematuria. Kidney Int 1996; 49:222–225.
17. Hebert LA, Betts JA, Sedmak DD, Cosio FG, Bay WH, Carlton S. Loin pain-hematuria syndrome associated with thin glomerular basement disease and hemorrhage into renal tubules. Kidney Int 1996; 49:168–173.

A History of Immunosuppressive Drugs in the Treatment of Inflammatory Bowel Disease:

Origins at The Mount Sinai Hospital

Burton I. Korelitz, M.D. and Daniel H. Present, M.D.

Abstract

If the cause of Crohn's disease and ulcerative colitis turns out to be some immunopathologic mechanism, many of the steps leading to such an understanding of their pathogenesis can be attributed to concepts that originated at The Mount Sinai Hospital. Perhaps immodestly, we can claim a role in the acceleration and the acceptance of these concepts; however, many contributions were made by others, including Moschkowitz, Klemperer, Otani, Crohn, Ginzburg, Oppenheimer, Marshak, and Janowitz. This does not mean that clinicians and researchers from other institutions did not contribute to this understanding.

As happens so often in medical history, elucidation of many disease processes are serendipitous. The concept of autoimmune diseases was introduced when we were house officers at Mount Sinai. The early days of transplant surgery soon followed along with the introduction by Hitchings and Elion of azathioprine to inhibit rejection. The concept of immunosuppression slowly evolved into possible treatment of any disease thought to be caused by autoimmunity, including those diseases of the bowel, seen so frequently at The Mount Sinai Hospital: ileitis, granulomatous colitis, ileocolitis, and ulcerative colitis.

Although most of the world called granulomatous disease of the bowel Crohn's disease, it was only after the deaths of Drs. Crohn, Ginzburg, and Oppenheimer that we accepted this single eponym. However, we will always pay tribute to all three Mount Sinai physicians who wrote the original paper that described the disease.

When we served as interns and residents, many years already had passed since the publication on regional ileitis by Crohn, Ginzburg, and Oppenheimer (1). Management remained primarily in the hands of the surgeons. Drug therapy had been limited to a variety of sulfonamides, and gastroenterologists were slow to acknowledge that sulfasalazine differed from the others and should be favored. Even so, sulfasalazine didn't work as well in ileitis as it did in ulcerative colitis. The morbidity and surgical incidence for both diseases remained high and the mortality rate, particularly for children with ulcerative colitis, was unacceptable (2).

The first edition of Dr. Crohn's book included data on course, complications, and treatment of ileitis. Corticosteroids, which became available soon thereafter, were not mentioned. Never before had such dramatic improvements been seen as those that followed treatment with corticosteroids and corticotropin. The outcome was influenced favorably so that surgery could be postponed and performed electively. The gastroenterologist inherited the primary responsibility

---

From the Lenox Hill Hospital, Department of Medicine, Division of Gastroenterology, and the Mount Sinai School of Medicine, Department of Medicine, Division of Gastroenterology, New York, New York. Address reprint requests to either: Burton I. Korelitz, M.D., 45 East 85th Street, New York, N.Y. 10028 or Daniel H. Present, M.D., 12 East 86th Street, New York, NY 10028.

for managing these patients and for clinical observation (3).

In retrospect this period was one of relative complacency, particularly with regard to the search for new drugs. The initial euphoria warranted by the obvious elimination of the inflammatory process by steroids yielded to the stark realization and disappointment that the effect was transient (4) and that the inflammation returned not only with discontinuation of the drug but even with dose reduction. Also the inflammation responded less dramatically with subsequent courses of treatment and, finally, corticosteroids were no more effective than placebo in maintaining remission (5).

In the 1960s the following factors rekindled an effort to provide new treatment for inflammatory bowel disease (IBD):

1. The causes remained unknown.
2. The concept of autoimmune diseases began to be appreciated (6, 7).
3. The question frequently was raised as to whether Crohn's disease and ulcerative colitis could be autoimmune diseases.
4. Studies demonstrating auto-antibodies in ulcerative colitis were being reported (8).
5. Azathioprine was shown to inhibit rejection of the donor's transplanted kidney (9).
6. Demonstrated toxicity due to corticosteroids was reported (10).
7. Winkelman and Brown (11) reported a favorable experience with nitrogen mustard in the treatment of both ulcerative colitis and regional enteritis.
8. Bean had successfully treated a patient who had ulcerative colitis with 6-mercaptopurine (12), a metabolic product of azathioprine, and was pursuing this early success (13).
9. Brooke et al. (14, 15) reported previously unheard of reversals of advanced Crohn's disease when patients were treated with azathioprine, including those with complex fistulas, after all surgical efforts had failed.

Although later epidemiologic data showed that the incidence of Crohn's disease was on the rise whereas the incidence of ulcerative colitis was stable, this was not yet apparent in the 1960s. One of us (BIK) had seen many patients who had ulcerative colitis with Banks and Zetzel at the Beth Israel Hospital in Boston (16). At Mount Sinai he was impressed with the many children with ulcerative colitis in whom corticosteroid therapy failed, who were physically and emotionally traumatized by the toxicity of these drugs, who required colectomy and ileostomy, and who died (3, 17). In view of this experience, in January 1967, it was decided to add 6-mercaptopurine (6-MP) to the treatment of patients with ulcerative colitis who had not responded favorably to corticosteroids and sulfasalazine treatment and who did not have an absolute indication for surgical intervention (18). The launching of this effort proved to be a major and all-consuming one. The co-worker in this study was Nathaniel Wisch, a hematologist, who favored 6-MP over azathioprine (azathioprine is converted to 6-MP *in vivo*) only because he had experience with its use in treating childhood leukemia. Furthermore, the Food and Drug Administration had approved 6-MP for use in the treatment of that disease. At the conclusion of this study, it was clear that 6-MP had no dramatic short-term influence on the course of the disease; moreover, transient nausea and vomiting occurred frequently. Consequently, 6-MP had no role in the treatment of fulminating ulcerative colitis or toxic megacolon. This finding confirmed those already demonstrated for azathioprine by Bowen et al. (19), who suggested that immunosuppressive drugs should not be used for the treatment of ulcerative colitis. Nonetheless, we had noted favorable results in a select group of patients which led us to adopt new criteria for instituting 6-MP therapy. Our choice of candidates included those with chronic prolonged ulcerative colitis, those who had an incomplete response to steroids, those who had complications from the steroid treatment, those who had contraindications to the use of steroids, and those in whom sulfasalazine had failed.

Perhaps the most important observation we made was that the real value of 6-MP could only be determined after long-term observation. We found that 13 of 14 patients either went into remission or improved significantly; only one remained incapacitated. Based on this study, and our later experience when the series had been expanded to 25 patients (20), a long-term, double-blind controlled study was planned.

Coincidentally, in the late 1960s the following events occurred:

1. One of us (DHP) completed his fellowship in gastroenterology at Mount Sinai. In addition to the teacher (BIK)-student (DHP) relationship, we became friends and found that we shared common interests for clinical investigation in the field of IBD. We were both enthusiastic about conducting a double blind trial of immunosuppressive therapy.
2. The Ileitis Foundation was launched by Irwin Rosenthal, William Modell, and our Chief of

Gastroenterology at Mount Sinai, Henry D. Janowitz. The primary goal of the Foundation was to encourage and support research in IBD, and DHP became its first Fellow and research grant recipient. The early participation of BIK was encouraged by the founders because of his accumulated experience with ulcerative colitis. Soon thereafter the name of the organization was changed to the National Foundation for Ileitis and Colitis.

3. Dr. Janowitz suggested that the protocol place more emphasis on Crohn's disease than on ulcerative colitis. It was already known that ulcerative colitis was predisposed to the development of cancer and there was some concern that immunosuppressive drugs might provoke an earlier onset of neoplasia. It was not yet known that Crohn's colitis also predisposed to the development of colon cancer. Also, we knew that a colectomy "cured" ulcerative colitis, but Crohn's disease recurred in most patients after surgical resection.

Crohn's Disease

In 1969 we launched our 2-year placebo-controlled, double-blind, crossover study of 6-MP in the treatment of refractory Crohn's disease (21). The study was conducted primarily through our private offices. Because there were no prior indices to measure the activity of Crohn's disease, we established our own for the controlled trial. These included "goals" of therapy for each case as might be done in the usual clinical management of Crohn's disease.

We felt then, as we do now, that Crohn's disease has many diverse presentations and that this "goal index" was the most accurate way to interpret clinical responses in patients with chronic disease. The most common goals were:

- elimination of corticosteroids
- closing of fistulas
- prevention of small bowel obstruction.

Elimination of the primary bowel symptoms alone was not an objective because steroid treatment temporarily accomplished this. Rather, symptoms had to be alleviated "after" stopping the steroids which indicated that 6-MP had been effective. It should be noted that although this study was initiated in 1969, it is still the only randomized controlled trial for treatment of fistulas in Crohn's disease. Although initiated after our study was started, the National Cooperative Crohn's Disease Study (NCCDS) was then being conducted at 14 other medical centers. The newly developed Crohn's Disease Activity Index (CDAI) was used as a measure of response to compare prednisone, sulfasalazine, azathioprine, and placebo (22, 23). This study concluded that azathioprine was ineffective (22). Based on the results of our own ongoing study, we and others submitted a critique of the NCCDS study pointing out its shortcomings (24, 25). Although the results of the NCCDS study were presented at the plenary session of the annual meeting of the American Gastroenterological Association (AGA) in 1980, our own abstract was not scheduled for presentation at that time. Rather, it was scheduled for presentation 1 day later and at a smaller forum. We encouraged clinical researchers to not yet disregard the use of immunosuppressive drugs in this disease and stated that physicians in private practice were capable of contributing significantly in the performance of randomized clinical trials. The basis of our concern regarding the outcome of the NCCDS study was the following:

1. The study terminated prematurely at 17 weeks, too soon for many patients to respond.
2. It eliminated azathioprine because it caused pancreatitis in a few instances. Consequently, there were insufficient entries assigned to azathioprine to reach statistical significance.
3. It eliminated steroids before starting the assigned drug or placebo, thereby enhancing the likelihood of exacerbating the underlying disease.
4. The dose of azathioprine used was relatively small.

In 1980 our own study was published and showed conclusively that 6-MP was effective in Crohn's disease (21). The following is a list of the most clear-cut features of the study.

1. The overall success rate was greater than 66%. Of the 39 patients who participated in each year of the 2-year crossover study in which either 6-MP or placebo was administered throughout one full year, the outcomes were significantly different statistically ($p < .0001$) (Table 1).
2. Thirty-three patients completed only one year and therefore received only one drug, either 6-MP or placebo; for the 19 receiving 6-MP, the rate of improvement was 79% as compared with 20% of 14 patients treated with placebo ($p < .05$) (Table 2).
3. To determine whether the outcome of the first year of trial influenced the outcome of the sec-

TABLE 1
Results in 39 Crossover Patients

Treatment	Number of patients	
	Improved[a]	Not improved
6-MP	26/39	13/39
Placebo	3/39	36/39

[a] Whereas 67% of the patients improved with 6-MP, only 8% improved with placebo. The difference is 59% with 95% confidence limits of 32 to 86% ($p < .0001$).

TABLE 2
Results in 33 Noncrossover Patients

Treatment	Number of patients	
	Improved[a]	Not improved
6-MP	15/19	4/19
Placebo	4/14	10/14

[a] Whereas 79% of the patients improved with 6-MP, only 29% improved with placebo. The difference is 50% with 95% confidence limits of 20 to 80% ($p < .05$).

ond year, a separate analysis was performed for 36 patients. A highly significant difference favored 6-MP with improvement in 67% versus 14% with placebo ($p < .0001$) (Table 3).

4. The study completely eliminated steroids in 55% of the patients or enabled dose reduction in 20%.
5. It promoted the healing of fistulas.
6. The study confirmed that the drug was slow acting (mean response time, 3.1 months) and in many (but not all) instances required continuation of steroids. Almost 20% of patients who would ultimately respond had not done so at 17 weeks, the time at which the NCCDS already had been completed.
7. Drug toxicity was modest and there was no early mortality.

More specific data on fistulas followed (26). 6-MP seemed to close fistulas in one third and meaningfully reduced discharge from fistulas in another third, with a mean response time of 3.5 months. 6-MP (and azathioprine) proved to be the first drug to have this influence on fistulas in Crohn's disease with the exception of metronidazole, which had a favorable effect only on perirectal fistulas. There has never been a controlled-trial confirmation of the efficacy of metronidazole on fistulas, but in our clinical experience many patients relapse with prolonged use of this antibiotic. Abdominal wall fistulas improved or closed in 10 of 12 (83%), rectovaginal fistulas in 4 of 6 (67%), and enteroenteric fistulas in 5 or 7 (71%). The perirectal abscesses and fistulas had sometimes led to diversionary surgery, such as a colostomy or ileostomy. The results in many cases had been disastrous, with recurrent Crohn's colitis in the stoma, diversion colitis in the rectum, and persistence of perirectal sepsis. A modified Parks' operation combined with 6-MP eradicated many perirectal abscesses and fistulas in those with recurrence after an incision and drainage procedure.

TABLE 3
Combined Results during First Year

Treatment	Number of patients	
	Improved[a]	Not improved
6-MP	26/36	10/36
Placebo	5/36	31/36

[a] Whereas 72% of patients improved with 6-MP, only 14% improved with placebo. The difference is 58% with 95% confidence limits of 40 to 77% ($p < .001$).

Our early observations noted that the use of immunosuppressive drugs apparently was more effective in the treatment of Crohn's disease when the colon (ileocolitis or colitis) was involved rather than only the small bowel. O'Donoghue et al. (27) had shown that remission, once established, was maintained by azathioprine in 79% of the patients at one year but not with placebo. Early observations after completion of our own controlled trial showed maintenance of remission in 19 out of 20 patients (95%) after a mean of 37 months on 6-MP. After stopping 6-MP therapy, 81% of the 32 patients who had been studied relapsed after a mean interval of 6 months. When 6-MP was reintroduced in 16 of these patients after relapse, improvement was noted within a mean response time of 1.5 months, considerably more rapid than had been the case for their first response (28).

Subsequently BIK and his colleagues at Lenox Hill reported a long-term experience with 6-MP in the treatment of Crohn's Disease (29). Again therapeutic goals were established for each of the 148 patients and an index of Crohn's disease activity was calculated pre- and post-therapy as well. The goals, as defined previously, were achieved in 68% of cases. Major successes included elimination of steroids (66%; $p < .001$) and healing of internal fistulas and abscesses with elimination or reduction in drainage and tenderness (64%; $p < .05$). In addition, drainage from perirectal fistulas and abscesses was eliminated or reduced as were the usually associated pain and tenderness (87% $p < .05$). Moreover, healing or marked improvement was noted in all cases of Crohn's disease involving the stomach and duodenum. Elective surgical resection was made much easier, if and when it had to be performed after 6-MP therapy, because the margins separating normal and diseased tissue were much better delineated. Despite these encouraging outcomes with 6-MP therapy, prevention of recurrent small bowel obstruction and disappearance of abdominal masses were noted in only 43% and 55% of the cases, respectively.

Ulcerative Colitis

After the early success by Bean (12, 13) in the treatment of ulcerative colitis with 6-MP, uncontrolled trials of immunosuppressives in ulcerative colitis reported response rates of 70 to 80%. There was less enthusiasm for initiating controlled trials in ulcerative colitis than in Crohn's disease due, in part, both to the risk of carcinoma of the colon and the knowledge that ulcerative colitis could be "cured" by colectomy. Furthermore, those who did undertake controlled trials reported disappointing results with one exception. In that study, it was noted that azathioprine permitted a significant reduction in steroid dosage (30).

Present et al. first reported in abstract form in 1988, a good to excellent response in 25 patients (73%). The mean time to the response was 2.3 months. A subsequent report from Mount Sinai (31) showed complete remission in 68 of 105 patients (65%) and partial remission in 25 (24%) with 12 (11%) failures. Although 6-MP was maintained, subsequent relapse was seen in 35% of the patients. In this group, steroids were required in 12% to restore remission. Fifteen patients stopped 6-MP electively and 13 relapsed after a median interval of 14 months. Thus, it was recognized that maintenance 6-MP was important after patients had responded favorably. Lobo et al. (32), in confirmation, have reported initial remission in 13 of 28 patients (46%). Eleven patients of 13 maintained their remission at 2 years.

Adler and Korelitz (33) reported the results of 87 patients refractory to steroids who had been treated with 6-MP. In 42 patients (48%), steroids could be eliminated with good control of symptoms after a mean treatment period of 2.5 months. The mean steroid-free period while on 6-MP was 10.9 months. Eleven cases (13%) demonstrated an intermediate but significant reduction in steroid usage.

Hawthorne et al. (34) reported the results of a randomized 1-year controlled withdrawal study of azathioprine versus placebo involving 67 patients with ulcerative colitis who had been taking azathioprine for 6 months or longer. In the 2-year interval, 59% of those switched to placebo relapsed in contrast to those maintained on azathioprine.

Despite the lack of controlled trials, in our experience with severe chronic ulcerative colitis, 6-MP has proven to be both safe and effective. We have been able to avoid urgent surgery in more than 75% of cases. In many cases, surgery has been recommended without making the patient aware of the efficacy of 6-MP. Perhaps to some physicians and surgeons this is justified by the advent of the ileal pouch anal anastomosis. This position, however, has been compromised by the lack of universal and consistent success with this procedure (mean technical failure rate of 6%) and the inability to predict which patients will do well and which patients will exchange one affliction for another (chronic pouchitis in 6–8%) (35–37).

Toxicity of Immunosuppressive Agents

Historical Aspects

Experience with 6-MP and azathioprine has shown that bone marrow suppression can occur and is dose related. Both leukopenia and thrombocytopenia have been reported. Hepatotoxicity is seen less frequently.

6-MP and azathioprine can cause chromosomal aberrations in animals and humans, but these are reversible in humans. The drugs are carcinogenic in animals at the dosages used and therefore may theoretically increase the risk of neoplasia in humans. Again the risk has been greater in transplant cases than in autoimmune disorders such as rheumatoid arthritis, perhaps because of the higher doses used.

Allopurinol, used to lower blood uric acid, may retard the catabolism of 6-MP and accentuate its toxicity. When the two drugs are used concurrently, the initial dose of 6-MP should be reduced by at least one third.

Toxicity to 6-MP in IBD has been compiled for 396 patients seen by the authors in their private practices (38). Four types of toxicity directly attributable to 6-MP have been identified. These are pancreatitis, bone marrow depression, idiosyncratic and/or hypersensitivity reactions, and drug-induced hepatitis.

Pancreatitis. Pancreatitis was observed in 12 (3%) patients with Crohn's disease and one with ulcerative colitis. The peak incidence was 21 days after starting the drug, and all but one case occurred within 32 days. Clinically each case was mild with symptoms of abdominal pain, nausea, vomiting, and elevation of serum amylase, all resolving quickly upon discontinuation of the drug. No hemorrhagic pancreatitis, hypocalcemia, pancreatic abscess, or pseudocyst formation was observed. Seven patients were later re-challenged with 6-MP or azathioprine and in each the symptoms recurred rapidly (39). Clinically the pancreatitis behaved like a sensitivity reaction, but attempts at desensitization were unsuccessful. Therefore, a single bout of 6-MP pancreatitis is considered a contraindication for future use, but a trial using small doses of azathioprine may be clinically warranted.

Bone Marrow Depression. In 2% of patients, bone marrow suppression sufficient to necessitate hospitalization occurred during the early years before experience was gained in blood count monitoring and individual dosage adjustment (38). In these eight patients the lowest white blood count ranged from 300 to 2500/mm^3. Fever was present in seven, and blood cultures were positive in three. Major septic complications were seen in one patient, but there were no deaths. Characteristically, the bone marrow depression was usually followed by prolonged remission of the Crohn's disease or ulcerative colitis. This observation has also been noted by others in relation to ulcerative colitis (40). Leukopenia was noted at some time in most patients. White blood counts above 3500/mm^3 were rarely associated with clinical problems and were managed by temporarily stopping the 6-MP and then resuming it at a lower dosage.

Subsequently, Connell et al. (41) reported the experience from St. Mark's in 739 patients with IBD treated with azathioprine. Only nine (1.2%) developed severe leukopenia (white blood cell count $< 2000/\text{mm}^3$); five of the nine patients had complications which resulted in two deaths. None of the deaths were predictable based on monitoring of the white blood cell count. Our now experience suggests that current methods of monitoring are adequate and that this severe complication is rare and may be modified by the availability of the colony-stimulating factor.

Idiosyncratic/Hypersensitivity Reactions. In addition to pancreatitis, other allergic type reactions occurred within the first 3 weeks of treatment with 6-MP in 2% of patients: drug fever was seen in eight, skin rash in two, arthralgia in one, and abdominal pain with normal serum amylase in one. We have also seen severe muscle pain soon after initiating treatment with 6-MP. Re-challenging with 6-MP or azathioprine produced similar reactions. Nevertheless, desensitization with gradually increasing doses has often allowed reinstitution of the drug without recurring adverse reaction (42). Whether these reactions are idiosyncratic or represent an allergic phenomenon is unknown, because no definitive immunologic hypersensitivity mechanism has been defined.

Drug-Induced Hepatitis. Eleven patients in this study developed a type of hepatitis, but in only one patient could it clearly be attributed to 6-MP. In this patient, who had severe ulcerative colitis that had remitted, results of liver function tests returned to normal after the drug was discontinued. These same tests became abnormal again after re-challenge. This is the only allergic type form of toxicity that seems to occur late in the course in most cases. Hepatitis is also unique in that re-challenge is often tolerated. The only instances of immediate toxicity occurred at a later date in a patient whose 6-MP was initiated at 3 mg/kg of body weight.

Infections

Almost every type of infection reported in the patient with IBD treated with 6-MP has also been reported in patients not treated with immunosuppressive drugs. This is especially true of the patients with Crohn's disease, in which infectious complications are frequent.

All of our patients receiving 6-MP who develop fever are instructed to discontinue the drug regardless of whether we think the fever is related to the underlying bowel disease, the drug, or an incidental illness such as an upper respiratory tract infection or a viral infection. One patient developed a liver abscess that required drainage. There had been no bone marrow suppression in this case.

Five patients developed pneumonia; all responded to antibiotics and cessation of 6-MP therapy. Some patients with pneumonia responded equally well to antibiotics while continuing 6-MP. Some patients reported frequent colds and upper respiratory tract infections and elected to discontinue 6-MP. It is difficult, however, to document any increased incidence of infection.

Herpes zoster occurred in eight patients while they were taking 6-MP. In all cases, the "shingles" took the usual course even though the dose of 6-MP was reduced by half. Many years later, one patient had a recurrence of herpes zoster on his trunk when he was again taking 6-MP and also suffered a 3-day syndrome consistent with encephalitis. There has been no residual defect during a 9-year follow-up, and the patient's ulcerative colitis remained in remission. Subsequent experience suggests there is no advantage to stopping the 6-MP during the course of the herpes zoster.

Neoplasms

The situation in regard to neoplasms is similar to that with infections. Neoplasms occur more frequently in patients with IBD whether or not they have received immunosuppressive medication. In Crohn's disease particularly, an increased risk of extraintestinal neoplasms has been recognized (43). An appraisal of neoplastic risk has been made during or after 6-MP therapy (38). In one instance the neoplasm was considered probably related to the 6-MP. This was a diffuse cerebral histiocytic lymphoma in a man with Crohn's disease who had taken 6-MP for approximately 9 months. The drug was discontinued 11 months before the onset of headaches and the discovery of the tumor. The tumor did not respond to therapy. Two more cases of cerebral lymphoma in patients with Crohn's disease treated with azathioprine have been reported subsequently (44).

Other neoplasms occurring in patients who had earlier received 6-MP included an islet cell carcinoma of the pancreas, carcinoma of the lung in a heavy cigarette smoker, one carcinoma of the breast, and a basal cell carcinoma. One patient with ulcerative colitis developed malignant melanoma; an excision with an axillary node dissection was curative. We have subsequently seen three patients who developed carcinoma of the breast who had previously received 6-MP.

Two carcinomas of the colon occurred in one patient with long-standing ulcerative colitis who had refused the recommendation of colectomy. In these cases, the carcinomas clearly anteceded the initiation of 6-MP. Subsequently one additional patient with long-standing left-sided ulcerative colitis, who had received 6-MP for 1 year, developed a carcinoma of the rectum. Benign adenomatous polyps have been encountered on sigmoidoscopy or colonscopy in patients receiving 6-MP; the polyps have been removed endoscopically and no recurrence has been encountered despite continuation of 6-MP.

Cancer of the colon could be a complication of particular concern because both ulcerative colitis and Crohn's colitis are diseases prone to the development of cancer even without immunosuppressive drugs. We reviewed our own experience and counted 34 patients with cancer of the colon complicating ulcerative colitis during a 20-year period; two had taken 6-MP, and 32 had not. What about cancer of the intestinal tract complicating Crohn's disease? We identified 17 patients in whom carcinoma of the ileum or colon developed; only two had been treated with 6-MP. Although these figures are not controlled and do not include a denominator, carcinoma seems to be far more common in Crohn's disease and ulcerative colitis independent of immunosuppressive therapy than associated with it.

Subsequently, Connell et al. (45) reported their data from St. Marks and concluded that the incidence of colorectal cancer and death from cancer in patients with IBD who had been treated with azathioprine is not greater than those patients who had not.

Nausea

Nausea is seen as a fairly common side effect during the first month of therapy. Occasionally the dose of 6-MP has to be temporarily reduced. Rarely does the drug need to be eliminated on this basis.

Mortality

With the single exception of the patient with cerebral lymphoma no deaths were attributable to 6-MP. Lymphomas of the brain (46) and other lymphomas (47) also have been reported in transplant patients on immunosuppressive therapy. Although it is most likely that this tumor was caused by 6-MP, two reports (48, 49) suggest an increased incidence of lymphomas in patients with Crohn's disease who are not taking immunosuppressives.

Considerations in Pregnancy

6-MP and azathioprine cause reversible chromosome damage in humans, but the true teratogenic effect of these drugs has not been studied in a systematic or controlled manner. On theoretical grounds, most obstetricians have advised that conception is contraindicated for both women and men receiving 6-MP. In chronically ill female patients we advise against pregnancy because of the persistent bowel activity. Therefore, before initiating 6-MP therapy, a female patient should have her serum tested for human chorionic gonadotropin level to eliminate the possibility of early pregnancy. All couples are instructed to practice a reliable method of birth control. The authors (BIK and DHP), however, differ with respect to the continued use of maintenance 6-MP therapy after suitable remission in patients who want to become pregnant.

When pregnancy becomes a priority, BIK recommends stopping 6-MP before discontinuing contraception. In a toxicity study (38), the authors found that 16 pregnancies had occurred. In ten patients, 6-MP had been terminated before contraception. In the remaining six patients, 6-MP had been continued for 3 to 4 weeks before the pregnancy was recognized. In three of the six patients, as well as the 10 who stopped 6-MP earlier, the pregnancies were continued. All 13 patients delivered full term. No abnormalities have been found in any of the 13 children. Additional data reported from St. Marks supports the conclusion that there is little danger to the mother or the fetus by taking azathioprine throughout pregnancy (50).

A recent extensive study of DHP's patients compared the outcomes of 240 live births in 155 patients with IBD who had stopped 6-MP before conception, had conceived while taking 6-MP, and then either stopped or continued the drug throughout the pregnancy to matched controls. There were no significant differences between the groups. It was concluded that 6-MP therapy either before conception or after conception and during pregnancy was not associated with prematurity, spontaneous abortion, congenital abnormalities, neonatal and childhood infections or neoplasia. These more recent data add support to the safety of 6-MP in pregnancy.

Cyclosporine

The marked difference between cyclosporine and 6-MP or azathioprine in the treatment of inflammatory bowel disease is that a favorable response, when seen, is more apt to be seen earlier after administration of the former drug, usually within 1 to 2 weeks rather than the 3 to 4 months for the latter drugs. Cyclosporine acts on T rather than B lymphocytes and inhibits their proliferation. The rapid onset of action is due to inhibiting the production of interleukin-2 and the consequent release of multiple cytokines.

The first controlled trial of oral cyclosporine (using 5 to 7 mg/kg/day) in patients chronically ill with Crohn's disease, performed by Brynskov et al. (51), showed a statistically significant response rate of 59% but, coincidentally, a 31% response for placebo. A subsequent controlled trial failed to confirm this response in chronically active patients, although slightly lower doses were used (52). Another controlled trial showed no difference in relapse rate with cyclosporine as compared to placebo using low doses of cyclosporine (less than 5 mg/kg/day) (53).

A new 12-month study (54) has been reported by European investigators using 5 mg/kg/day, and no major difference has been found when comparing cyclosporine plus low-dose steroid versus low-dose steroid alone.

A different technique was initiated at Mount Sinai where Lichtiger and colleagues used "intravenous" cyclosporine to treat Crohn's disease as well as ulcerative colitis (55). At an intravenous dose of 4 mg/kg/day for 14 days, which was then followed by oral doses of 6 to 8 mg/kg/day, a favorable response rate was noted in 88% of the patients. Closure of Crohn's disease fistulas occurred in 44% of these patients and improvement was noted in another 44%. Two thirds maintained improvement in the chronic phase and steroids could be discontinued in 75%. Hanauer and Smith (56) showed a similar response rate using this technique. Thus, oral cyclosporine seems to be less effective than intravenous cyclosporine, which may be due to malabsorption of the drug and/or the need for higher doses in inflammatory bowel disease patients with IBD.

The most promising data on cyclosporine comes from Present and Lichtiger and their colleagues at Mount Sinai who treated patients with ulcerative colitis via the intravenous route at 4 mg/kg/day for 14 days after a 7 to 10-day trial of parenteral steroids had been unsuccessful (57). A mean response time of 6.4 days was observed and colectomy was avoided in 75%. A subsequent controlled trial has confirmed the efficacy of intravenous cyclosporine under these circumstances (58).

Further observation has shown that oral cyclosporine has been successful for 6 months in 69% (59). This major long-term difference between the response in Crohn's disease and ulcerative colitis might be attributed to the malabsorption so prevalent in the former and not in the latter.

Toxicity is of great concern when using cyclosporine, particularly intravenously. The foremost concern is nephrotoxicity which fortunately can be minimized by carefully monitoring the blood urea nitrogen (BUN) and creatinine, keeping the cyclosporine blood level in a therapeutic range, and avoiding nonsteroidal anti-inflammatory agents and other drugs which may enhance the levels of cyclosporine and its toxicity. The risk of lymphoma and carcinoma is of concern, but in view of the doses and duration used for the treatment of IBD, cyclosporine's use has not been forstalled. Other complications, such as tremors, hirsutism, gingival hyperplasia, paresthesia, headaches, and hypertension, although common, are almost always transient.

Although the experience with cyclosporine in

IBD is limited, a preliminary conclusion that intravenous administration will prove to be an important adjunct in the management and treatment of very sick patients with ulcerative colitis in whom treatment with intravenous steroids has failed is justified. In the experience of BIK the number of patients with this combination will be small because the response with intravenous steroids is usually sufficient to buy time to establish successful treatment with chronic phase medications such as 5-aminosalicylic acid products and 6-MP. The major risk in using intravenous cyclosporine when intravenous steroids have failed is waiting too long. Although the waiting period probably should be no longer than 7 to 12 days, some patients are already too sick to await any hoped for response to cyclosporine. Patients so toxic are vulnerable to pneumonia, sepsis, and death when they might have been saved by a timely colectomy. On the other hand, should there be time for a trial of intravenous cyclosporine, its success will allow for introduction of the slow-acting immunosuppressive drug 6-MP which may then maintain remission for many years thereafter. This decision, of course, has to be based on a finely attuned clinical evaluation of the individual patient.

The Present and the Future

Of all the immunosuppressive drugs now available, 6-MP and azathioprine have had the most enduring role and continue to show the most promise for the future. Despite the disadvantage of its longer delay to initial response once introduced, 6-MP has been consistently effective in two thirds of patients with otherwise refractory Crohn's disease and those with ulcerative colitis. Azathioprine has also been effective for the same groups of patients. Because 6-MP is a metabolic by-product of azathioprine, presumably weight for weight its favorable effect is not as great as the purer compound, 6-MP. Nevertheless, this has never been submitted to any controlled trial.

What is more important is the answer to the question as to why favorable results of treatment are consistently about two thirds in patients with IBD. Why is it not 100%? What accounts for the other one third? Perhaps the answer in Crohn's disease lies with the pathology, particularly of the small bowel in which repeated attacks of inflammation result in fibrosis which to date has been irreversible. After all, success with 6-MP in preventing small bowel obstruction has been consistently less than for other indications, such as elimination of steroids and closure of fistulas.

Perhaps the answer lies in the role of inducing leukopenia with 6-MP or azathioprine. In retrospect those patients who did develop leukopenia had greater success than those who did not (60). Furthermore, the success came earlier and was longer lasting, and those who did develop leukopenia had no clinical bone marrow depression. This can be attributed in part to what we have learned about careful monitoring. Our approach is to perform complete blood counts weekly for the first 3 to 4 weeks, and then if the white blood count does not fall and clinically there is no indication for increasing the dose of 6-MP, the time between blood counts can be extended. If, however, at any time the symptoms recur, and the white blood count permits, the dose of 6-MP should be raised. At this point, monitoring should be on done weekly for 3 weeks. Another time for particularly careful monitoring is when steroids have been reduced or are about to be eliminated. Because the steroids had served to raise the white blood count, this stimulus may be lost and the patient may be vulnerable. In all cases the patient must call after the blood count to receive new directions as to dose and when the next count should be done. Platelets also should be counted because thrombocytopenia is seen infrequently during 6-MP therapy.

Another question is why leukopenia develops so easily in one patient but not in another. Although some patients tolerate as much as 200 mg a day without apparent effect on the white blood count, others tolerate only tiny doses such as 12.5 mg per week, or even less. When leukopenia is first recognized in these patients they are almost always doing better than those without leukopenia, but when the dose has to be further and further reduced, the value of leukopenia is replaced by loss of overall drug efficacy.

In conclusion, physicians with roots at Mount Sinai have played a major role in evaluating the use of immunosuppressive drugs in the treatment of inflammatory bowel disease. We anticipate a future with increased productivity in the academic tradition of this great institution (61–63).

References

1. Crohn BB, Ginzburg L, Oppenheimer GB. Regional ileitis: A pathological and clinical entity. JAMA 1932; 99: 1323–1329.
2. Korelitz BI, Gribetz D, Danziger I. The prognosis of ulcerative colitis with onset in childhood. I. Presteroid era. Ann Intern Med 1962; 57:582–591.
3. Korelitz BI, Lindner AE. The influence of corticotrophin and adrenal steroids in the course of ulcerative colitis. Gastroenterology 1964; 46:671–679.
4. Spencer SA, Kirsner JB, Mlynaryk P, Reed PI, Palmer WL. Immediate and prolonged therapeutic effects of

corticotrophin and adrenal steroids in ulcerative colitis: Observations in 340 cases of periods up to ten years. Gastroenterology 1962; 42:113–129.

5. Lennard-Jones JE, Misiewicz JT, Connell AM, et al. Prednisone as maintenance treatment for ulcerative colitis. Lancet 1965; 1:188–189.
6. Mackay IR, Burnet FM. Autoimmune disease: Pathogenesis, chemistry, and therapy. Springfield, IL: Charles C. Thomas, 1963: 237–239.
7. Skinner MD, Schwartz RS. Medical progress: Immunosuppressive therapy. N Engl J Med 1972; 287:221–227.
8. Broberger O. Immunologic studies in ulcerative colitis. Gastroenterology 1964; 47:229–240.
9. Elion GB, Hitchings GH. Azathioprine. In: Handbook of Experimental Pharmacology. New York: Springer Verlag. 1975:404–425.
10. Sparberg M, Kirsner JB. Long term corticosteroid therapy for regional enteritis. Am J Dig Dis 1966; 11:865–880.
11. Winkelman EJ, Brown GH. Nitrogen mustard in the treatment of chronic ulcerative colitis and regional enteritis. Clevel Clin Q 1965; 32:165–174.
12. Bean RHD. The treatment of chronic ulcerative colitis with 6-mercaptopurine. Med J Aust 1962; 2:592–593.
13. Bean RHD. Treatment of ulcerative colitis with antimetabolites. Br Med J 1966; 1:1081–1084.
14. Brooke BN, Hoffmann DC, Swarbrick ET. Azathioprine for Crohn's disease. Lancet 1969; 2:612–614.
15. Brooke BN, Javett SL, Davison OW. Further experience with azathioprine for Crohn's disease. Lancet 1970; 2:1050–1053.
16. Banks BM, Korelitz BI, Zetzel L. The course of nonspecific ulcerative colitis. Reviewing twenty years experience and late results. Gastroenterology 1957; 32:983–1012.
17. Korelitz BI, Gribetz D. The prognosis of ulcerative colitis with onset in childhood. II. The steroid era. Ann Inter Med 1962; 57:592–597.
18. Korelitz BI, Wisch N. Long term therapy of ulcerative colitis with 6-mercaptopurine: A personal series. Am J Dig Dis 1972; 17:111–118.
19. Bowen GE, Irons GV Jr., Rhodes JD, Kirsner JB. Early experience with azathioprine in ulcerative colitis. JAMA 1966; 195:166–170.
20. Korelitz BI, Glass JL, Wisch N. Long-term immunosuppressive therapy of ulcerative colitis. Continuation of a personal series. Am J Dig Dis 1973; 18:317–322.
21. Present DH, Korelitz BI, Wisch N, Glass JL, Sachar DB, Pasternack BS. Treatment of Crohn's disease with 6-mercaptopurine. A long-term, randomized, double blind study. N Engl J Med 1980; 302:981–987.
22. Summers RWL, Switz DM, Sessions JT Jr, et al. National Cooperate Crohn's Disease Study: Results of drug treatment. Gastroenterology 1979; 77:847–869.
23. Best WR, Becktel JM, Singleton JW, Kern F. Development of a Crohn's Disease Activity Index. Gastroenterology 1976; 70:433–444.
24. Goldstein F, Thornton JJ, Abramson J. Comments on National Cooperation Crohn's Disease Study (NCCDS). Gastroenterology 1980; 78:1647–1648.
25. Korelitz BI, Present DH. Shortcomings of the National Cooperative Crohn's Disease Study: In the exclusion of azathioprine without adequate trial. Gastroenterology 1981; 80:193–194.
26. Korelitz BI, Present DH. Favorable effect of 6 mercaptopurine in fistulas of Crohn's disease. Dig Dis Sci 1985; 30:58–64.
27. O'Donoghue DP, Dawson AM, Powell-Tuck J, et al. Double blind withdrawal trial of azathioprine as maintenance treatment of Crohn's disease. Lancet 1978; 2:955–957.
28. Korelitz BI. Therapy of inflammatory bowel disease including use of immunosuppressive agents. Clin Gastroenterol 1980; 9:331–349.
29. Korelitz BI, Adler DJ, Mendelsohn RA, Sacknoff AL. Longterm experience with 6-mercaptopurine in the treatment of Crohn's disease. Am J Gastroenterol 1993; 88:1198–1205.
30. Rosenberg JL, Wall AJ, Levin B, et al. A controlled trial of azathioprine in the management of ulcerative colitis. Dig Dis 1975; 20:721–726.
31. George J, Pou R, Bodian C, et al. 6-Mercaptopurine (6-MP) is effective in chronic ulcerative colitis but how long do you use it? Gastroenterology 1994; 106:A686.
32. Lobo AJ, Foster PN, Burke DA, et al. The role of azathioprine in the management of ulcerative colitis. Dis Colon Rectum 1990; 33:374–377.
33. Adler DJ, Korelitz BI. The therapeutic efficacy of 6 mercaptopurine in refractory ulcerative colitis. Am J Gastroenterol 1990; 85:717–722.
34. Hawthorne AB, Logan RFA, Hawkey CJ, et al. Randomized controlled trial of azathioprine withdrawal in ulcerative colitis. Br Med J 1992; 305:20–22.
35. Gemlo BT, Wong WD, Rotherberger DA, Goldberg SM. Ileal pouch anal anastomosis. Arch Surg 1992; 127: 784–786.
36. Salemans JMJI, Nagengat FM, Lubbers ESC, Kuijper's JH. Post-operative and long-term results of ileal pouch anal anastomosis for ulcerative colitis and familial polyposis coli. Dig Dis Sci 1992; 37:1882–1889.
37. Setti-Carraro P, Ritchie K, Wilkinson KH, et al. The first 10 years experience of restorative proctocolectomy for ulcerative colitis. Gut 1994; 35:1070–1075.
38. Present DH, Meltzer SJ, Krumholz M, et al. 6-mercaptopurine in the management of inflammatory bowel disease: Short-term and long-term toxicity. Ann Intern Med 1989; 3:641–649.
39. Haber CJ, Meltzer SJ, Present DH, et al. Natural course of pancreatitis caused by 6-mercaptopurine in the treatment of inflammatory bowel disease. Gastroenterology 1986; 91:982–986.
40. Burke DA, Dixon MF, Axon ATR. Ulcerative colitis. Prolonged remission following azathioprine-induced pancreatitis. J Clin Gastroenterol 1989; 11:327–330.
41. Connell WR, Kamm MA, Lennard-Jones JE, et al. Twenty-seven years experience in 739 patients of bone marrow toxicity from azathioprine in inflammatory bowel disease. Gastroenterology 1992; 102:A609.
42. Korelitz BI, Present DH, Rubin PH, et al. Desensitization to sulfasalazine after hypersensitivity reactions in patients with inflammatory bowel disease. J Clin Gastroenterol 1984; 6:27–31.
43. Greenstein AJ, Gennuso R, Sachar DB, et al. Extraintestinal cancers in inflammatory bowel disease. Cancer 1985; 56:2914–2921.
44. Lemann M, Bonhomme P, Bitoun A, et al. Traitement de la maladie de Crohn par l'azathioprine ou la 6-mercaptopurine. Étude retrospective chez 126 maladies. Gastroenterol Clin Biol 1990; 14:548–554.
45. Connell WR, Kinlen LJ, Ritchie JK, et al. Cancer risks from azathioprine in inflammatory bowel disease. Gastroenterology 1994; 106:A667.
46. Schenk SA, Penn I. De-novo brain tumors in renal transplant recipients. Lancet 1971; 1:983–986.
47. Penn I, Hammond W, Brettschneider L, et al. Malignant

lymphoma in the transplantation patient. Transplant Proc 1969; 1:106–112.

48. Glick SN, Teplick SK, Goodman LR, et al. The development of lymphoma in patient with Crohn's disease. Radiology 1984; 153:337–339.
49. Greenstein AJ, Mullin GE, Strauchen JA, et al. Lymphoma in inflammatory bowel disease. Cancer 1992; 69:1119–1123.
50. Alstead EM, Ritchie JK, Lennard-Jones JE, et al. Safety of azathioprine in pregnancy in inflammatory bowel disease. Gastroenterology 1990; 90:443–446.
51. Brynskov J, Freund L, Rasmussen SM, et al. A placebo controlled, double-blind randomized trial of cyclosporine, therapy in active Crohn's disease. N Engl J Med 1985; 321:845–850.
52. Jewell DP, Lennard-Jones JE and the Cyclosporine Study Group of Great Britain and Ireland. Oral cyclosporine for chronic active Crohn's disease: A multicenter controlled trial. Eur J Gastroenterol Hepatol 1994; 6:499–505
53. Feagan BG, McDonald JWD, Rochon J, et al. Low dose cyclosporine for the treatment of Crohn's disease. N Engl J Med 1994; 330:1846–1851.
54. Stange EF, Modigliani R, Pena AS, et al. European trial of cyclosporine in chronic active Crohn's disease: A 12-month study. Gastroenterology 1995; 109:774–782.
55. Present DH, Lichtiger S. Efficacy of cyclosporine in treatment of fistula of crohn's disease. Dig Dis Sci 1994; 39:374–380.
56. Hanauer SB, Smith MB. Rapid closure of Crohn's disease fistulas with continuous intravenous cyclosporine. Am J Gastroenterol 1993; 88:646–649.
57. Lichtiger S, Present DH. Preliminary report: Cyclosporine in treatment of severe active ulcerative colitis. Lancet 1990; 1:16–19.
58. Lichtiger S, Present DH, Kornbluth A, et al. Cyclosporine in severe ulcerative colitis refractory to steroid therapy. N Engl J Med 1994; 330:1841–1845.
59. Kornbluth A, Lichtiger S, Present DH, Hanauer S. Long-term results of oral cyclosporine in patients with severe ulcerative colitis: A double blind, randomized multicenter trial (Abstract). Gastroenterology 1994; 106: A714.
60. Colonna T, Korelitz BI. The role of leukopenia in the 6-mercaptopurine induced remission of refractory Crohn's Disease. Am J Gastroenterol 1994; 89:362–366.
61. Present DH. Cyclosporine and other immunosuppressive agents: Current and future role in the treatment of inflammatory bowel disease. Am J Gastroenterol 1993; 88:627–630.
62. Korelitz BI. My thoughts on using cyclosporine in the treatment of inflammatory bowel disease (Editorial). Am J Gastroenterol 1995; 90:2095–2086.
63. Korelitz BI, Present DH. Methotrexate for Crohn's Disease. N Engl J Med 1995; 333:600–601.

Pathogenesis of Inflammatory Bowel Disease:

Evolving Concepts

Lloyd Mayer, M.D., Xian Yang Yio, M.D., Annica Lin, B.A., and Asit Panja, M.D.

Diseases marked by chronic inflammation remain obstinate in their responses to therapy and elusive in their definition of pathogenesis. Such is the case for the inflammatory bowel diseases (IBD), ulcerative colitis and Crohn's disease, whose clinical course is characterized by remissions and exacerbations and whose underlying pathogenetic mechanisms are still unknown. Our knowledge is hampered by the fact that inflammation is a normal response to pathogens but in these disorders inflammation is not regulated. As in any chronic inflammatory disease, alterations in immunologic function are usually invoked but rarely definitively shown to be responsible for disease expression. This is more marked in IBD in which immunologic dysfunction is occurring in an organ in which the rules are distinct from those regulating systemic immunity. It is this latter concept that has developed over the past several years. Investigators, attempting to fit the dogma of systemic immunoregulation into the gut, have become aware of distinct cell populations and novel mechanisms for handling the usual trigger of any immune response, antigen. The growing concept in IBD is that there are obvious defects in mucosal immune responses that occur in genetically susceptible individuals by potentially ubiquitous environmental stimuli (1, 2).

Phenomena Unique to Mucosal Immunity

Some of the interesting features that separate the mucosal from the systemic immune system probably relate to the challenges that exist in the different environments (**Table**). The intestinal lymphocytes are juxtaposed to a veritable morass of antigens, yet their responses are meager at best (3). This is best evidenced by the lack of good oral vaccines. With the exception of polio and more recently *Haemophilus influenzae* B, most orally administered vaccines result in no responses. Although this may relate to the breakdown of antigen in the gut into nonimmunogenic fragments, noninvasive bacteria that remain intact rarely result in active immunity. This phenomenon may be linked to the entity of oral tolerance (4–8). In animal models, feeding of antigen results in antigen-specific systemic immunosuppression mediated by $CD8^+$ T cells (7, 8). The tolerance induced is generally within both the cellular and humoral compartments of the immune response (although greater on the cellular side). Recent evidence shows that oral tolerance can be generated in man but its effects appear to be predominantly T cell oriented (9). Induction of tolerance and active suppression of responses to specific antigen has been used in the therapy of a variety of autoimmune disorders in animals and man (10–14). Although quite successful in rat and mouse, the evidence for its success in outbred humans remains to be seen.

A phenomenon linked to oral tolerance is the controlled inflammation in the gut. The lamina propria is filled with activated memory cells that remain in a relatively constant state of activation

---

This work was supported by U.S. Public Health Service grants AI 23504, AI 24671, and DK 44156.

Address reprint requests to Lloyd Mayer, M.D., The Mount Sinai School of Medicine, Division of Clinical Immunology, Department of Medicine, Box 1089, One Gustave L. Levy Place, New York, NY 10029.

TABLE
Phenomena Unique to Mucosal Immunity

Oral tolerance
Controlled inflammation
Secretory IgA secretion in the setting of variable systemic suppression
Unique cell populations—intraepithelial lymphocytes, lamina propria lymphocytes

(15, 16). Despite large numbers of luminal organisms and dietary antigens, the cell numbers and activation state remain relatively constant. This appears to be yet another example of the generally suppressed tone of the gut. Last, the intestine is marked by unique cell populations. Lymphocytes reside within the epithelial space (IEL) at a ratio of 1:10 to 1:100 lymphocytes to epithelial cells. These features are unique as they are virtually all T cells, without access to locally conventional antigen-presenting cells (APCs), e.g., macrophages, dendritic cells with which to interact. The overwhelming majority of these cells express CD8 (17–20). In the mouse there is some evidence that these cells may be derived extrathymically with the intestinal epithelium acting as a thymic equivalent (21). What the IELs do and why they are predominantly of the CD8 subset has not been defined, although theories regarding a surveillance role and regulation of epithelial integrity abound (22). The lamina propria lymphocytes (LPL), as mentioned earlier, are a chronically activated cell population which do not necessarily follow the rules of conventional immunity. These cells are activated memory cells that home to the gut via special homing receptors but once present are difficult to activate through the antigen receptor (23, 24). They may be local cytokine factories, but who or what they are regulating remains unknown.

Mucosal Immune System

As alluded to above, immune responses in the gut differ greatly from those elicited systemically. By the nature of its location, juxtaposed with the external environment, this makes intuitive sense. The gut is the largest lymphoid organ in the body housing 10^{10} IgA-producing lymphocytes/meter, predominantly in the form of secretory sIgA (25). These immunocompetent cells are separated from a sea of antigens by a one-cell-thick layer of epithelium that functions under normal conditions to absorb nutrients and water and to secrete a protective mucous barrier. Several nonimmunologic mechanisms protect the host from luminal pathogens (see the Table). Most importantly, these include a thick mucous barrier, preventing attachment of microorganisms to the epithelium, and a variety of chemicals (acid) and enzymes (proteases) that destroy pathogens and render ingested antigens nonantigenic (reducing large macromolecules into di- and tripeptides). One cannot underestimate the importance of these factors to significantly limit antigen exposure by nonimmune mechanisms of exclusion.

Several antigens bypass these barriers and can access the mucosal lymphoid tissue, however. Here too there are differences between mucosal and systemic immunity. Antigen in the form of macromolecules (dietary) or organisms can be sampled in the gut by one of three pathways: through specialized epithelium or M cells overlying organized lymphoid tissue or Peyer's patches; between epithelial cells (paracellular); or through the epithelial cell (transcellular) (26) **(Fig. 1)**. Studies to date support the fact that antigens binding to and trafficking through M cells result in the initiation of a positive immune response (27–29). This includes activation of helper T cells ($CD4^+$) and induction of antigen-specific IgA-producing B lymphocytes that will eventually home back to other mucosal surfaces (e.g., lung, genitourinary tract, mmamary gland). Trafficking of antigen between epithelial cells (paracellular) occurs only under conditions where there is existing inflammation and tight junctions joining adjacent epithelial cells (preventing entry of molecules in the normal state) are disrupted allowing for entry of luminal pathogens into the underlying lamina propria (30). The third pathway, transcellular, has been invoked recently as a potential mechanism for antigen sampling in the normal bowel (31–34). Predigested peptides as well as some bacteria and viruses can enter the intestinal epithelial cell (IEC) and translocate to the basolateral surface *in vitro* and in some settings *in vivo* (35). This latter scenario is important because it may define an important immunoregulatory pathway for the maintenance of intestinal homeostasis. *In vitro* studies in various species have documented that intestinal epithelial cells express class II molecules (HLA-DR, DP, DQ in man, I-A, I-E in mouse) on their surface that serve as regulatory elements in T cell antigen-presenting cell interactions (31–33, 36). Co-culture of IEC and T cells in the presence of soluble protein antigens can trigger T-cell proliferation and cytokine production (31–33). However, in contrast to conventional APCs where such an interaction would result in the activation of $CD4^+$ helper T cells,

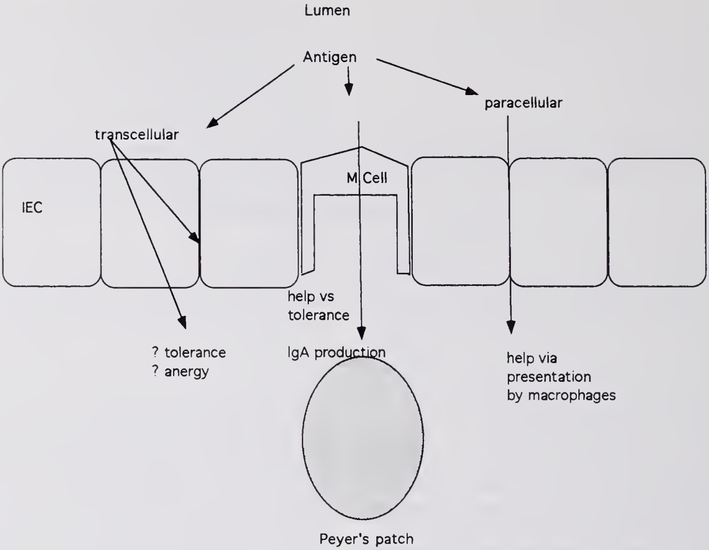


Fig. 1. Pathways of antigen trafficking across the intestinal epithelium.

IEC:T cell co-cultures result in the expansion of $CD8^+$ T cells (31, 32).

Here is an excellent example of where the systemic and mucosal immune systems differ. In systemic immunity the dogma states that the CD4 molecule on T cells binds to major histocompatibility complex (MHC) class II molecules on APCs. Antigens taken up exogenously are processed in endosomal compartments where they associate with class II molecules (37). The class II MHC/peptide complex is then expressed on the cell surface where it can be recognized by a specific T-cell antigen receptor. CD4 binds to a nonpolymorphic domain of class II enhancing the association of the T cell and the APC as well as bringing in a series of kinases and substrates to participate in T-cell activation (38). The difference between mucosal and systemic immunity is that despite the expression of class II molecules on the IEC membrane, CD8 and not $CD4^+$ T cell proliferation occurs in the normal state (31, 32).

On the surface this finding is difficult to reconcile and several models have been proposed to explain this dichotomy. Kaiserlian and her colleagues have suggested that the class II molecules expressed on IEC are distinct from those on conventional APCs (39). This does not appear to be the case in humans and rats (40). A second hypothesis relates to the possibility that there may be alternate molecules expressed on IEC that might preferentially promote activation of $CD8^+$ cells bearing this molecule. Likely candidates here would be nonclassical class I molecules that have been shown to be expressed on murine and human IEC and may be responsible in part for *in vitro* $CD8^+$ T-cell proliferation induced by IEC (41, 42).

The question remains as to how this activation occurs. Studies attempting to define which surface molecular interactions are important in this system have demonstrated that neither classical class I nor class II molecules mediate $CD8^+$ T-cell:IEC interactions (43). Furthermore despite the expression of class II molecules, antibodies to CD4 fail to inhibit T-cell proliferation in these cultures (43). However, CD8 does appear to be involved and this involvement extends to the activation of the CD8-associated src-like tyrosine kinase p56lck (44). Activation of this kinase is necessary for $CD8^+$ T-cell activation induced by IEC. To define the molecule responsible for such activation of $CD8^+$ T cells, we have generated a series of monoclonal antibodies (mAb) against normal epithelium. Two mAbs recognizing a 180-kilodalton glycoprotein (gp180) were found to be potent inhibitors of $CD8^+$ T-cell proliferation and one of these mAbs inhibits the activation of CD8-associated p56lck (43). This glycoprotein is expressed throughout the normal gastrointestinal tract from duodenum to rectum as well as on thymic epithelium and syncytiotrophoblast cells (the maternal/fetal interface) (Yio and Mayer, in preparation). This antigen is not found on skin epithelium, mammary epithelial cells, and only one of the two mAbs recognizes a structure on airway epithelium. The findings in the thymus and placenta are of interest because, if this molecule is involved in the generation of suppression, these are two sites where such a process would be important.

What then does this 180-kilodalton glycoprotein represent? We have shown that purified gp180 binds to CD8 and activates p56lck (Yio and Mayer, submitted). Thus it could represent either a novel restriction element or more likely on adhesion molecule with binding specificity for CD8. Antigens trafficking through the epithelium might associate with some restriction element but be focused onto CD8$^+$ T cells by virtue of the expression of gp180. Clearly our understanding of the molecule awaits its molecular characterization and cloning; however its biology still of considerable interest.

Immoregulatory Defects in IBD

How does this relate to the inflammatory process in IBD? As alluded to earlier there is a growing consensus as to mechanisms involved in the pathogenesis of IBD. Defects in normal mucosal immunoregulation appear to be a prominent feature. Studies from many laboratories have centered around the finding that there is aberrant activation of CD4$^+$ T cells in the tissue of patients with both Crohn's disease and ulcerative colitis (45, 46). This is evidenced by the expression of a variety of activation markers on these cells within the lamina propria and the spillover of these activated cells into the peripheral blood. CD4$^+$ lamina propria T cells in IBD express higher levels of interleukin (IL) 2R, class II molecules, transferrin receptor, and activation markers CD69 and CD71 (45–48). The results of this activation are clear. Activated CD4$^+$ T cells are the major cytokine secretors. It is of interest to note that Romagnani (49) subcategorized Th cells into two groups, T_h1 and T_h2 (**Fig. 2**). Th1 cells secrete IL2, γ-interferon (γ-IFN), and tumor-necrotizing factor (TNF) and promote classical cellular immune or DTH responses. These types of reactions result in macrophage activation and granuloma formation, the type of histology associated with Crohn's disease (50). T_h2 cells, in contrast, secrete cytokines that promote antibody responses IL4, IL5, IL6, and IL10. Antibody-mediated tissue injury with complement activation and autoantibody production is charac-

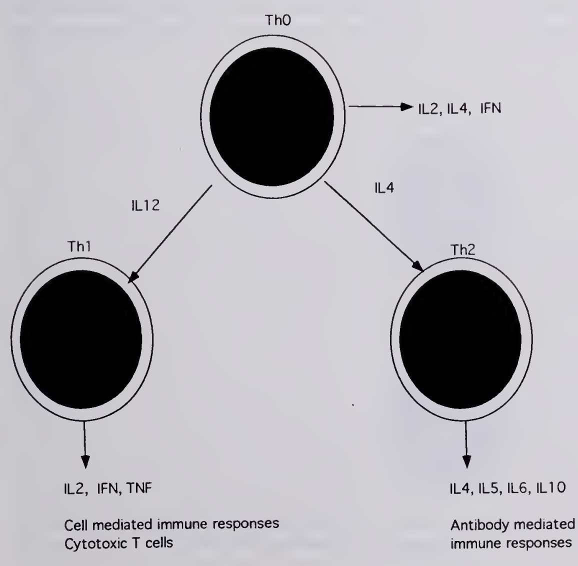


Fig. 2. T helper subsets: cytokine production.

teristic of what is seen in ulcerative colitis (51). Numerous studies have suggested indirectly that such a T_h1/T_h2 dichotomy exists in IBD. Scott and his colleagues (52) reported that lamina propria B cells from patients with ulcerative colitis produced more IgG than normal lamina propria B cells and that the majority of this IgG was of the IgG_1 subclass. In Crohn's disease there was an increase in IgG as well, except that in this setting IgG_2 was the major contributor (52). The increase in IgG_1 speaks to the presence of IL4, a potent stimulus for IgG_1 and IgE production. Interestingly, patients with ulcerative colitis have a greater incidence of atopic disease (a T_h2-mediated, IgE-mediated disorder) and a higher familial incidence of atopy, which supports the concept that there may be a genetic predisposition toward the activation of T_h2 cells (53). Autoantibodies have been described in ulcerative colitis more than in Crohn's disease with antibodies to perinuclear antigens (ANCA) and epithelial cells (antitropomyosin-anti-40kda) (54, 55). Thus ulcerative colitis may be a T_h2-mediated disease. In Crohn's disease the production of IgG_2 is consistent with the presence of γ-IFN, a known regulator of IgG_2 production, as well as an inducer of macrophage activation. Thus Crohn's disease with its classic delayed type hypersensitivity like histologic appearance may be a T_h1-mediated disease. Mullin et al. (56) have given direct evidence for this. Using mRNA extracted from mucosal biopsies they have demonstrated that by quantitative polymerase chain reaction, IL2 and γ-IFN mRNA are increased in patients with Crohn's disease, whereas IL5 and IL4 mRNA are increased in ulcerative colitis (56, 57). This dichotomy in cytokine production occurred only at sites of active disease, suggesting that local activation of the $CD4^+$ T-cell subsets is occurring.

The next logical question is what is driving this activation? If epithelial cells are playing a role in the activation of $CD8^+$ suppressor T cells in the normal state, maybe these cells are defective in IBD. The concept of defective suppression in chronic inflammatory diseases is not novel and in fact has been invoked in IBD in several experimental systems (58, 59). Our laboratory has shown that there is a defect in the activation of $CD8^+$ T cells by IEC isolated from patients with IBD (60) (**Fig. 3**). Importantly this defect appears to be present whether the IEC are isolated from areas of active inflammation or from areas without histologic disease. Therefore, there may be an inherent defect in IBD epithelium that results in a failure of suppressor cell generation. This finding appears to be unique to IBD because IEC derived from other inflammatory states (diverticulitis, ischemic colitis, radiation colitis) are quite capable of activating $CD8^+$ T cells despite the enhanced expression of class II molecules on these cells (60).

Because gp180 may be involved in $CD8^+$ T-cell activation in the normal state, defects in production or expression of this glycoprotein may

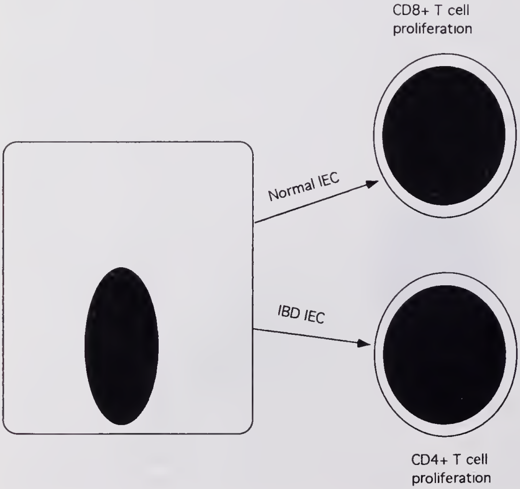


Fig. 3. Schematic of activation of T cells by intestinal epithelial cells.

be present in IBD. Preliminary studies suggest that this may be the case. gp180 is not expressed on IEC derived from patients with Crohn's disease and is expressed in a patchy distribution in patients with ulcerative colitis (Lin et al., in preparation). Similar to the findings with intact IEC, defects in gp180 expression are found in both actively inflamed and uninflamed IEC. Failure to express gp180 could result in unopposed activation of $CD4^+$ T cells by class II MHC-bearing enterocytes, sampling greater amounts of luminal antigen. Although this is an attractive hypothesis, it should be recognized that IEC have not been shown to function as APC *in vivo*. Furthermore the early lesions of Crohn's disease overlie Peyer's patches, suggesting entry of antigen through the M cell and not through epithelium (61). In ulcerative colitis the frank ulceration allows for free access of luminal antigens to the immunocompetent cells in the lamina propria. Here tissue macrophages might be capable of presenting antigen to $CD4^+$ cells sustaining an active inflammatory state. Thus there are alternate models to explain $CD4^+$ T-cell activation in IBD.

Regardless, newer therapies for IBD have focused on the inappropriate activation of $CD4^+$ T cells. Intravenous cyclosporin A, a potent inhibitor of cytokine production by activated $CD4^+$ T cells, is a powerful therapy for acute active ulcerative colitis (62). Therapy with anti-CD4 mAbs as well as lymphapheresis have been effective at ameliorating disease, although only over the short term (63, 64). The development of AIDS with $CD4^+$ T-cell cytopenia induced remission in several patients with Crohn's disease but was not associated with remission in ulcerative colitis (65). This difference in outcome with Crohn's disease versus ulcerative colitis may relate to the predominant T_h2 response seen in HIV disease which would inhibit T_h1 T-cell cytokine production (66). Therapeutic trials with anticytokine antibodies have been initiated but are too preliminary to judge. However, cytokine gene knockout mice have resulted in the expression of intestinal inflammation (67, 68). Interestingly knockout of the gene encoding IL10 (a T_h2 cytokine that inhibits T_h1 clones) results in a picture consistent with Crohn's disease, essentially an unopposed Th1 state. A second model, transferring Th1-like cells into an immunodeficient (SCID) mouse also promoted a Crohn's like lesion (69). These studies support the concept that the mucosal immune system is tightly regulated and that disruption of this regulation is tolerated poorly in the gut. Such disruption can be better compensated for in the more antigenic pristine environment of the systemic immune system. IBD should, therefore, be viewed as an immunodisregulated state; a phenomenon that is genetically predetermined in search of an environmental trigger.

References

1. Kirsner JB. Genetic aspects of inflammatory bowel disease. Clin Gastroenterol 1973; 2:557–575.
2. Yang H, McElree C, Roth MP, et al. Familial empirical risks for inflammatory bowel disease: Differences between Jews and non-Jews. Gut 1993; 34(4):517–524.
3. Zeitz M, Quinn TC, Graeff AS, James SP. Mucosal T cells provide helper function but do not proliferate when stimulated by specific antigen in lymphogranuloma venerum proctitis in nonhuman primates. Gastroenterology 1988; 94:353–366.
4. Challacombe SJ, Tomasi TB, Jr. Systemic tolerance and secretory immunity after oral immunization. J Exp Med 1980; 152:1459–1472.
5. Mattingly JA, Eardley DD, Kemp JD, Gershon RK. Induction of suppressor cells in rat spleen: influence of microbial stimulation. J Immunol 1979; 122:787–790.
6. Miller SD, Hanson DG. Inhibition of specific immune responses by feeding protein antigens. IV. Evidence for tolerance and specific active suppression of cell-mediated immune responses to ovalbumin. J Immunol 1979; 123:2344–2350.
7. Mowat AM. The role of antigen recognition and suppressor cell in mice with oral tolerance to ovalbumin. Immunology 1985; 56:253–000.
8. Mowat AM. The regulation of immune responses to dietary protein antigens. Immunol Today 1987; 8(3):93–98.
9. Husby S, Mestecky J, Moldoveanu Z, et al. Oral tolerance in humans. T cell but not B cell tolerance after antigen feeding. J Immunol 1994; 152(9):4663–4670.
10. Whitacre CC, Gienapp IE, Orosz CG, Bitar DM. Oral tolerance in experimental autoimmune encephalomyelitis. III. Evidence for clonal anergy. J Immunol 1991; 147(7):2155–2163.
11. Barinaga M. Treating arthritis with tolerance. Science 1993; 261:1669–1670.
12. Okumura S, McIntosh K, Drachman DB. Oral administration of acetylcholine receptor: effects on experimental myasthenia gravis. Ann Neurol. 1994; 36(5):704–713.
13. Weiner HL, Mackin GA, Matsui M, et al. Double-blind pilot trial of oral tolerization with myelin antigens in multiple sclerosis. Science 1993; 259:1321–1324.
14. Trentham DE, Dynesius-Trentham RA, Orav EJ, et al. Effects of oral administration of type II collagen on rheumatoid arthritis. Science 1993; 261:1727–1730.
15. James SP, Zeitz M, Kanof M, Kwan WC. Intestinal lymphocyte populations and mechanisms of cell-mediated immunity. Immunol Allergy Clin North Am 1988; 8: 369–391.
16. Zeitz M, Green WC, Peffer NJ, et al. Lymphocyte isolated from the intestinal lamina propria of normal nonhuman primates have increased expression of genes associated with T-cell activation. Gastroenterology 1988; 94:647–655.
17. Mowat AM. Human intraepithelial lymphocytes. Semin Immunopathol 1990; 12:165–190.
18. Ebert EC. Proliferative responses of human intraepithe-

lial lymphocytes to various T-cell stimuli. Gastroenterology 1989; 97:1372–1381.

19. Halstensen TS, Scott H, Brandtzaeg P. Human CD8+ intraepithelial T lymphocytes are mainly CD45RA-RB+ and show increased co-expression of CD45RO in celiac disease. Eur J Immunol 1990; 20:1825–1830.
20. Croitoru K, Ernst PB. Leukocytes in the intestinal epithelium: an unusual immunological compartment revisited. Regul Immunol 1992; 4(2):63–69.
21. Guy-Grand D, Malassis-Seris M, Briottet C, Vassalli P. Cytotoxic differentiation of mouse gut thymodependent and independent intraepithelial lymphocytes is induced locally: Correlation between functional assays, presence of perforin and granzyme transcripts and cytoplasmic granules. J Exp Med 1991; 173(6):1549–1552.
22. Hershberg R, Eghtesady P, Sydora B, et al. Expression of the thymus leukemia antigen in mouse intestinal epithelium. Proc Natl Acad Sci USA 1990; 87(24):9727–9731.
23. James SP, Kwan WC, Sneller MC. T cells in inductive and effector compartments of the intestinal compartments of the intestinal mucosal immune system of nonhuman primates differ in lymphokine mRNA expression, lymphokine utilization, and regulatory function. J Immunol 1990; 144:1251–1256.
24. Targan SR, Deem RL, Liu M, Wang S, Nel A. Definition of a lamina propria T cell responsive state. Enhanced cytokine responsiveness of T cells stimulated through the CD2 pathway. J Immunol 1995; 154(2):664–675.
25. Mestecky J, McGhee JR, Elson CO. Intestinal IgA system. Immunol Allergy Clin North Am 1988; 8:349–368.
26. Mayer L. Antigen presentation in the intestine. In: Targan SR, Shanahan F, eds. Immunology and immunopathology of the liver and gastrointestinal tract. New York: Igaku-Shoin, 1990.
27. Wolf JL, Bye WA. The membranous epithelial (M) cell and the mucosal immune system. Annu Rev Med 1984; 35:95–112.
28. Owen RL, Jones AL. Epithelial cell specialization within human Peyer's patches: An ultrastructural study of intestinal lymphoid follicles. Gastroenterology 1974; 66: 189–203.
29. Sicinki PJ, Rowinski JB, Warchol A, et al. Poliovirus type I enters the human host through intestinal M cells. Gastroenterology 1990; 98:56–58.
30. Madara JC, Staffork J. Interferon γ directly affects barrier function of cultured intestinal epithelial monolayers. J Clin Invest 1989; 83:724–727.
31. Mayer L, Shlien R. Evidence for function of Ia molecules on gut epithelium cells in man. J Exp Med 1987; 166: 1471–1483.
32. Bland PW, Warren LG. Antigen presentation by epithelial cells of the rat small intestine. II. Selective induction of suppressor T cells. Immunology 1986; 58:9–14.
33. Kaiserlian D, Vidal K, Revillard JP. Murine enterocytes can present soluble antigen to specific class II restricted $CD4^+$ T cells. Eur J Immunol 1989; 19:1513–1516.
34. Bland PW, Warren LG. Antigen presentation by epithelial cells of the rat small intestine. I. Kinetics, antigen specificity and blocking by anti-Ia antisera. Immunology 1986; 58:1–7.
35. Brandeis JM, Sayegh MH, Gallon L, et al. Rat intestinal epithelial cells present major histocompatibility complex allopeptides to primed T cells. Gastroenterology 1994; 107(5):1537–1542.
36. Wiman K, Curman B, Forsum U, et al. Occurrence of Ia antigens on tissues of non-lymphoid origin. Nature 1978; 276:711–713.
37. Harding CV, Unanue ER, Slot JW, et al. Functional and ultrastructural evidence for intracellular formation of major histocompatibility complex class II-peptide complexes during antigen process. Proc Natl Acad Sci USA 1990; 87:5553–5557.
38. Veillette A, Abraham N, Caron L, Davidson D. The lymphocyte-specific tyrosine protein kinase p56lck. Semin Immunol 1991; 3:143–152.
39. Kaiserlian D, Nicolas JF, Revillard JP. Constitutive expression of Ia molecules by murine epithelial cells: A comparison between keratinocytes and enterocytes. J Invest Dermatol 1990; 94(3):385–386.
40. Mayer L, Eisenhardt D, Salomon P, et al. Expression of class II molecules on intestinal epithelial cells in man: Differences between normal and inflammatory bowel disease. Gastroenterology 1990; 100:3–12.
41. Bleicher PA, Balk SP, Hagen SJ, et al. Expression of murine CD1 on gastrointestinal epithelium. Science 1990; 250:679–682.
42. Blumberg RS, Terhorst C, Bleicher P, et al. Expression of a nonpolymorphic MHC class I-like molecule, CD1d, by human intestinal epithelial cells. J Immunol 1991; 147: 2518–2524.
43. Mayer L, Siden E, Becker S, Eisenhardt D. Antigen handling in the intestine-mediated by normal enterocytes. In: MacDonald TT, Challacombe SJ, Bland PW, Stokes CR, Heatley RV, Mowat AM, eds. Advances in Mucosal Immunology. Lancaster, UK: Kluwer Academic Publishers, 1990.
44. Li Y, Panja A, Mayer L. Epithelial cells activate T cells through the src-like tyrosine kinase, p56lck. Gastroenterology 1991; 100:A593.
45. Fais S, Pallone F, Squarcia O, Boirvant M, Pozilli P. T cell early activation antigens expressed by peripheral lymphocytes in Crohn's disease. J Clin Immunol 1985; 16: 75–76.
46. Targan SR, Kagnoff MF, Brogan MD, Shanahan F. Immunologic mechanisms in intestinal diseases. Ann Intern Med 1987; 106:853–870.
47. Mueller C, Knoflach P, Zielinski CC. T-cell activation in Crohn's disease: Increased levels of soluble interleukin-2 receptor in serum and in serum supernatants of stimulated peripheral blood mononuclear cells. Gastroenterology 1990; 98:639–646.
48. Schurmann G, Betzler M, Post S, Herfarth C, Meuer S. Soluble interleukin-2 receptor, interleukin-6 and interleukin-1 beta in patients with Crohn's disease and ulcerative colitis: preoperative levels and postoperative changes of serum concentrations. Digestion 1992; 51(1): 51–59.
49. Romagnani S. Induction of TH1 and TH2 responses: a key role for the 'natural' immune response? Immunol Today 1992; 13:379–381.
50. Riddell RH. Pathology of idiopathic inflammatory bowel disease. In: Kirsner JB, Shorter RG, eds. Inflammatory Bowel Disease. Philadelphia, PA: Lea and Febiger, 1988:329–350.
51. Halstensen TS, Mollnes TE, Garred P, et al. Epithelial deposition of immunoglobulin G1 and activated complement (C3b and terminal complement complex) in ulcerative colitis. Gastroenterology 1990; 98:1264–1271.
52. Scott MG, Nahm MH, Macke K, et al. Spontaneous secretion of IgG subclasses by intestinal mononuclear cells: Differences between ulcerative colitis, Crohn's disease, and controls. Clin Exp Immunol 1986; 66:209–215.

53. Gilat T, Hacohen D, Lilos P, Langman MJS. Childhood factors in ulcerative colitis and Crohn's disease: An international cooperative study. Scand J Gastroenterol 1987; 22:1009–1024.
54. Yang H, Rotter JI, Toyoda H, et al. Ulcerative colitis: a genetically heterogeneous disorder defined by genetic (HLA class II) and subclinical (antineutrophil cytoplasmic antibodies) markers. J Clin Invest 1993; 92(2): 1080–1084.
55. Takahashi F, Das KM. Isolation and characterization of a colonic autoantigen specifically recognized by colon tissue-bound immunoglobulin G from idiopathic ulcerative colitis. J Clin Invest 1985; 76:311–318.
56. Mullin GE, Lazenby AJ, Harris ML, et al. Increased interleukin-2 messenger RNA in the intestinal mucosal lesions of Crohn's disease but not ulcerative colitis. Gastroenterology 1992; 102(5):1620–1627.
57. Mullin GE. Expression of Th2 cytokines in the lamina propria in patients with IBD. Inflamm Bowel Dis, in press, 1996.
58. Hodgson HJ, Wands JR, Isselbacher KJ. Decreased suppressor cell activity in inflammatory bowel disease. Clin Exp Immunol 1978; 32:451–458.
59. Kelleher D, Murphy A, Whelan CA, et al. Defective suppression in the autologous mixed lymphocyte reaction in patients with Crohn's disease. Gut 1989; 30:839–844.
60. Mayer L, Eisenhardt D. Lack of induction of suppressor T cells by intestinal epithelial cells from patients with inflammatory bowel disease. J Clin Invest 1990; 86(4): 1255–1260.
61. Morson BC. The early histological lesion of Crohn's disease. Proc R Soc Med 1972; 65:71–72.
62. Lichtiger S, Present DH, Kornbluth A, et al. Cyclosporine in severe ulcerative colitis refractory to steroid therapy. N Engl J Med 1994; 330(26):1841–1845.
63. Deusch K, Mauthe B, Reiter C, Reithmuller G, Classen M II. CD4-antibody treatment of inflammatory bowel disease: One year follow up. Gastroenterology 1993; 104:A691.
64. Bicks R, Groshart K. T-lymphocyte apheresis (TLA) treatment in chronic active Crohn's disease. Gastroenterology 1993; 104:A669.
65. Greenwald B, James SP. Long-term HIV infection with Crohn's disease. Am J Gastroenterol 1995; 90(1):167–168.
66. Romagnani S, Del Prete G, Manetti R, et al. Role of TH1/TH2 cytokines in HIV infection. Immunol Rev 1994; 140:73–92.
67. Kuhn R, Lohler J, Rennick D, et al. Interleukin-10-deficient mice develop chronic enterocolitis. Cell 1993; 75(2):263–274.
68. Sadlack B, Merz H, Schorle H, et al. Ulcerative colitis-like disease in mice with a disrupted interleukin-2 gene. Cell 1993; 75(2):253–261.
69. Powrie F, Correa-Oliveira R, Mauze S, Coffman RL. Regulatory interactions between CD45RBhigh and CD45RBlow CD4+ T cells are important for the balance between protective and pathogenic cell-mediated immunity. J Exp Med 1994; 179(2):589–600.

Calcification of Elastic Fibers in Pseudoxanthoma Elasticum

Sharada Truter, Ph.D., Jodi Rosenbaum-Fiedler, M.D., Allen Sapadin, M.D., and Mark Lebwohl, M.D.

Abstract

Pseudoxanthoma elasticum is an inherited disorder of connective tissue that is characterized by calcification of elastic fibers with associated abnormalities of the skin, ocular, and cardiovascular systems. The genetic defect causing pseudoxanthoma elasticum has not been determined and the diagnosis relies on clinical recognition of skin lesions and histologic demonstration of calcification of elastic fibers that are clumped and fragmented in the dermis. The role of connective tissue components in the etiology of pseudoxanthoma elasticum is reviewed and includes changes in collagen, elastin, glycosaminoglycans, fibronectin, microfibrillar proteins, modifying enzymes of extracellular matrix proteins, fibroblasts, and fibrillins.

PSEUDOXANTHOMA ELASTICUM IS A RARE, heritable disorder of elastic fibers that is found in all population groups throughout the world. Pseudoxanthoma elasticum was described initially by Rigal (1) as a lipid disorder. In 1896, Darier (2) assigned the name "pseudoxanthoma elasticum" to the xanthoma-like plaques found in the skin flexures. The disease is characterized by symptoms of accelerated aging, including loss of vision, accelerated cardiovascular disease, and cutaneous lesions that cause teenage skin to look prematurely wrinkled. The largest series of patients in the world is followed up at the Mount Sinai Medical Center where approximately 500 patients with the disease have been seen. This article will review clinical, histologic, and pathophysiologic aspects of pseudoxanthoma elasticum, including contributions made at the Mount Sinai Medical Center.

Pseudoxanthoma elasticum is inherited through both autosomal recessive and autosomal dominant pathways (3–5). The estimated prevalence of pseudoxanthoma elasticum is thought to be approximately 1 in 100,000, but because we do not have a precise laboratory marker for the disease, its exact prevalence is unknown.

The primary etiology of pseudoxanthoma elasticum has been a matter of controversy for many years. Lack of familial cases in the sporadic and autosomal recessive cases of pseudoxanthoma elasticum and variable expression in autosomal dominant families make it difficult to perform genetic linkage studies. Substantial evidence supports a major role for alterations of elastic fibers in the pathogenesis of the disease. One of the most striking histologic features of pseudoxanthoma elasticum is calcification of elastic fibers in the middle and deep dermis. Ultrastructural and light microscopic abnormalities of both collagen and elastin (6) have been described in debating which of these is responsible for the primary defect in pseudoxanthoma elasticum. Very little attention has been focused on the other extracellular matrix components associated with skin development.

The disease mainly manifests itself within the skin, cardiovascular system, and the eye. Multiple organ systems are involved due to clumping and fragmentation of elastic fibers within the lesional areas. Cutaneous manifestations are characterized by the presence of yellow

---

From the Department of Dermatology, Mount Sinai School of Medicine, New York, New York. Address requests for reprints to Dr. Sharada Truter, Department of Dermatology, Mount Sinai School of Medicine, Box 1047, One Gustave Levy Place, New York, NY 10029.

papules and plaques resembling that of "plucked chicken-skin" or "cobblestones" with redundant folds of skin in flexural regions of the neck (Figure), axillae, antecubital and popliteal fossae, and periumbilical areas. Calcification of the internal elastic laminae of arteries can lead to accelerated cardiovascular disease (7), and patients can die of myocardial infarctions at an early age. Cracking of the calcified blood vessels can lead to bleeding, and cardiac valvular involvement is common (8). The cardiovascular manifestations range from changes resembling arteriosclerosis in the medium-sized arteries to hypertension, congestive heart failure, and stroke. A significant association with mitral valve prolapse was first demonstrated at the Mount Sinai Medical Center where it was shown that more than two thirds of patients are affected by this cardiac valvular abnormality (8).

In the eye, Bruch's membrane becomes calcified with resultant angioid streaks, retinal hemorrhages, and eventual loss of central vision. Pseudoxanthoma elasticum exhibits the Koebner phenomenon, meaning that skin lesions have a predilection for scars. With this in mind, Mount Sinai investigators showed that pseudoxanthoma elasticum can be demonstrated on biopsy of normal-appearing scars in patients who have angioid streaks but do not have skin lesions of the disease (9). Because there is no serologic marker for pseudoxanthoma elasticum, and the cutaneous lesions that are relied on for diagnosis may be absent, it is sometimes difficult to demonstrate that a patient has the disease. Even the classic ocular finding, angioid streaks, is absent in some patients (10) and is not specific for pseudoxanthoma elasticum. Many cases that have been thought incorrectly to be inherited in an autosomal recessive fashion may in fact be inherited in an autosomal dominant manner. Because the condition is not always detected in relatives with mild skin changes, autosomal dominant inheritance (11), which has been demonstrated in some families, may therefore be more common than is generally appreciated.

The Cause of Pseudoxanthoma Elasticum

Old Theories

Fibroblasts. Several skin biopsies derived from patients affected with pseudoxanthoma elasticum showed fibroblasts with extremely dilated endoplasmic reticulum, containing a dense granular material (12). The endoplasmic reticulum sometimes appeared to be empty with few dispersed grains or filaments.

Small cysts were also observed between the laminae of the nuclear envelope containing these grains and filaments. A close relationship between lesional fibroblasts and thin long nonbanded filaments exist. These thin long filaments were contiguous with the dilated endoplasmic reticulum of the pseudoxanthoma elasticum cells. Pinocytosis was also observed, as well as calcification of the thin long filaments and a "thready" material. Normal and abnormal appearing elastic fibers were associated with these thin long fila-

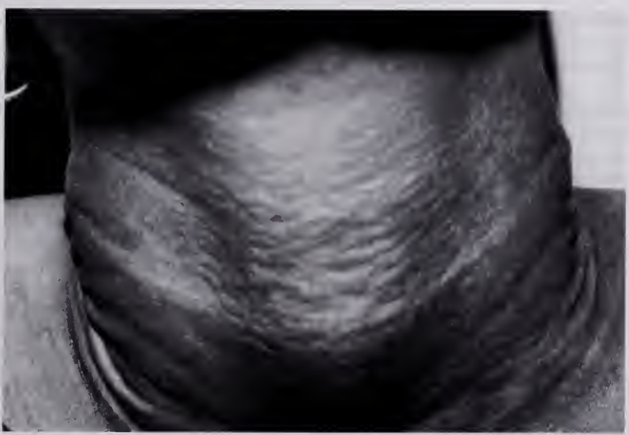


Fig. Redundant folds of the skin in the neck of a 20-year-old female patient with pseudoxanthoma elasticum.

ments and were found to contain calcified material. The degeneration of the elastic fibers appeared to vary according to the age of the patient. As a rule, even though all fibroblasts contained numerous ribosomes, the dilated cystic reticulum had few or no ribosomes and contained thin filaments or dispersed grains. The disappearance of ribosomes from the fibroblasts of the lesions may be evidence for an abnormal protein. The thin parallel filaments with a periodicity resembling that of collagen suggest that the precursor form of collagen fibrils might be associated with pseudoxanthoma elasticum.

Collagen. Using indirect immunofluorescence with antibodies against type III procollagen, an increase in type III collagen in affected areas of the dermis has been demonstrated (13).

Electron microscopic examination of the collagen fibers in pseudoxanthoma elasticum showed irregularly arranged normal or twisted collagen fibrils with increased diameters (11, 12). Collagen changes in pseudoxanthoma elasticum resembled those of other genetic diseases of connective tissue such as hyalinosis cutis et mucosae (14), tuberous sclerosis (15), and dermatosparaxis (16).

Elastic Fibers. Electron microscopy and immunoelectron microscopy have demonstrated abnormalities in a large proportion of the elastic fibers in lesional pseudoxanthoma elasticum skin, although a few normal fibers have been observed. At the Mount Sinai Medical Center, similar abnormalities were also found in nonlesional skin. Irregularly shaped holes were present in the center of many of the elastic fibers which were empty, with some containing a slightly electron-dense granular material. A distinctive ring of electron-dense material presumed to be calcium was identified in some elastic fibers.

Glycosaminoglycans. Martinez-Hernandez and Huffer (17) suggested that cells from patients with pseudoxanthoma elasticum synthesize an abnormal polyanionic substance, probably a glycosaminoglycan, that is deposited on elastic fibers and attracts calcium. The alteration and increase of elastic fibers were postulated to be a secondary event. Affected pseudoxanthoma elasticum dermis contains increased amounts of glucosamine (18), hyaluronic acid (19), and chondroitin sulfate B (20). Gordon et al. (21) suggested that pseudoxanthoma elasticum fibroblasts may produce an abnormal serine protease, thereby degrading the high molecular weight proteoglycans to lower molecular weights, enabling them to localize within the elastic fiber and initiate mineralization. Kobayasi et al. (22) have shown that the thin filaments with knobs found in the round holes in the elastic fiber matrix represent acid glycosaminoglycans.

Most of the studies that focused on the involvement of glycosaminoglycans with pseudoxanthoma elasticum were based on lesional skin in a small number of patients. Studies at the Mount Sinai Medical Center (23) examined glycosaminoglycans in nonlesional skin to elucidate the changes that precede and possibly lead to the calcification of elastic fibers. A marked increase of glycosaminoglycans in both lesional and nonlesional skin was demonstrated. Both hyaluronic acid and dermatan sulfate were found to have increased more in the lesional than in nonlesional skin, which contained more glycosaminoglycans than skin from healthy control subjects. The results suggest that the deposition of glycosaminoglycans precedes the histologic changes in pseudoxanthoma elasticum as well as the calcification of the elastic fibers in this disorder. Elevated levels of hyaluronic acid and dermatan sulfate were also observed in the urine of some patients with pseudoxanthoma elasticum (24). Some of the changes that occur in pseudoxanthoma elasticum elastic fibers can be reproduced by injecting glycosaminoglycans into rabbit dermis (25).

Modifying Enzymes. Protocollagen proline hydroxylase, the enzyme hydroxylating peptide-bound proline to collagen hydroxyproline, is increased in pseudoxanthoma elasticum (26). However, the ratio of hydroxyproline to protein is decreased significantly in pseudoxanthoma elasticum. The decreased ratio may be a result of increased collagenolysis or may be partly reflective of an increase in elastin and/or nonfibrous proteins in relation to collagen, which is in agreement with observations made by Smith et al. (18).

Fibronectin. A possible role for fibronectin in pseudoxanthoma elasticum has been explored. Indirect immunofluorescence has shown that lesional pseudoxanthoma elasticum skin contains increased amounts of fibronectin (27). The region of the dermis above the calcified area showed a decrease in fibronectin staining, and the pattern in the upper dermis resembled that of normal skin. The distribution pattern of the fibronectin in nonlesional skin did not differ from that of control skin. On immunoelectronmicroscopy with ferritin-tagged antibodies to fibronectin, this substance was localized to elastic microfibrils.

New Theories

Microfibrillar Proteins. Ultrastructural and biochemical analysis of elastic fibers in the extracellular matrix of tissues revealed a 10- to 12-nm microfibrillar component surrounding the abundant amorphous elastin component (28). During the process of elastogenesis, elastin is secreted into the extracellular matrix, as a 72-kDa tropoelastin precursor (29). In the extracellular matrix, the enzyme lysyl oxidase modifies the precursor to become highly cross-linked into a rubber-like network. On recent electron microscopic analysis, the 10-nm microfibrils appeared as thread-like filaments, regularly beaded in longitudinal sections and tubular in cross-section (30). The discovery of the recently isolated 260-kDa glycoproteins, the microfibrils fibrillin-1 and fibrillin-2, shed new light on the presence of the previously observed thin long microfibrils associated with the disease. The multiple fibrillin molecules align in a parallel head-to-tail fashion to form the major component of the microfibrils. Fibrillins contain multiple epidermal growth factor-like repeats that have been associated with calcium binding through the hydroxylation of asparagine or the aspartic acid located in the epidermal growth factor-like repeat. Other microfibril components have been identified including the 31-kDa and 78-kDa microfibril-associated glycoprotein (MAGP), the enzyme lysyl oxidase, the 34-kDa glycoprotein with aminooxidase activity, the 58-kDa associated microfibril protein (AMP), and emilin.

Elastin is synthesized by embryonic and rapidly growing tissues and cells derived from them. The prolyl residues in tropoelastin become hydroxylated post-transcriptionally. It was found that ascorbic acid improves the degree of hydroxylation that appears to interfere with the process of elastin fibrillogenesis and stabilization. The increased levels of hydroxyproline in elastic fibers of patients with Marfan syndrome seems to compromise the tensile strength of their aortas. Crucial to the function of elastic fibers is the high degree of cross-linking of tropoelastin by lysyl oxidase.

The process of elastin fibrillogenesis takes place at unique sites located at infoldings of the cell surface. The microfibrils are found grouped in small bundles near the plasma membrane directing the morphogenesis of the elastic fibers by acting as a scaffold on which the elastin is deposited. It is suggested that the microfibrils align the tropoelastin molecules such that the cross-linking regions are juxtaposed before oxidation by lysyl oxidase can occur. The majority of microfibrils are progressively displaced to the outer surfaces of the fiber, a position they retain in the mature tissue. Fibrillin-2, a fetal protein, is thought of as the microfibril that becomes embedded within the amorphous elastin structure during the process of morphogenesis. The carboxyl-terminal end of the tropoelastin molecule mediates the interactions to the microfibrils and is cleaved from the tropoelastin after it serves an alignment function. Other researchers have detected the carboxyl-terminal sequences in insoluble elastin, suggesting that tropoelastin does not undergo cleavage before cross-linking.

Fibrillins. Pseudoxanthoma elasticum shares some clinical features with Marfan syndrome (31, 32) in which defects in fibrillin-1 have been identified (33, 34). Unlike fibrillin-1, fibrillin-2 is sparse or absent in adult skin but richer in elastic tissues during fetal development (35–37). Thus, abnormalities of fibrillin-2, a fetal protein, can result in clinical disease later in life even though the abnormal protein is not identifiable in adult tissues. It has been suggested that the fibrillin-2 protein is either rapidly degraded or is masked during elastogenesis, and fibrillin-2 is postulated to act as a scaffold upon which newly formed elastin is deposited (12). Structural organization of the fibrillins reveals multiple epidermal growth factor-like repeats that bind calcium to stabilize the linear conformation of the molecule. Genetic mutations that affect these epidermal growth factor-like moieties may increase their affinity for calcium (38–41), potentially resulting in calcification of elastic tissue. Indirect immunofluorescence studies with fibrillin-1 antibodies on dermal fibroblasts derived from patients with Marfan syndrome documented a consistent decreased deposition of fibrillin-1 (42–45). The reduced pattern of expression of fibrillin-1 suggested that this was the candidate gene product in Marfan syndrome. Defects in fibrillin-1 are now known to cause the variable and pleiotropic manifestations of Marfan syndrome. Mutational analysis of the fibrillin-1 coding sequence has revealed more than 30 mutations that cause Marfan syndrome. Recent studies also revealed that the fibrillin-2 protein is the candidate gene product in the Marfan syndrome-related disorder, congenital contractural arachnodactyly (CCA) (33, 46, 47). Mutations found in epidermal growth factor-like motifs in fibrillin-2 are the cause of the CCA phenotype (48).

Although elastic fibers become calcified in patients with pseudoxanthoma elasticum, linkage to the tropoelastin gene has been excluded (49). In addition, no correlation was found by indirect immunofluorescence with fibrillin-1 secretion by dermal fibroblasts derived from pseudoxanthoma elasticum patients (50). By culturing adult dermal fibroblasts in an artificial skin grown on nylon meshes (51, 52) we have been able to demonstrate the presence of fibrillin-2 by indirect immunofluorescence. Computerized quantification of fluorescence shows a marked reduction of fibrillin-2 deposition by pseudoxanthoma elasticum dermal fibroblasts grown on the meshes compared to normal fibroblasts (53). Fibrillin-2 is thus implicated as a candidate gene for pseudoxanthoma elasticum.

Summary

In conclusion, pseudoxanthoma elasticum is a complex inherited disorder that profoundly affects elastic tissue. Complications affect many organ systems, including the heart, eye, skin, and blood vessels. Although the precise cause of pseudoxanthoma elasticum has not yet been determined, calcification of elastic tissue with deposition of glycosaminoglycans formally occurs in patients with this disorder. A possible primary role for fibrillin-2, a fetal protein, has been suggested as an etiology for this disease.

References

1. Rigal D. Observations pour sevir a l'histoire dela cheloide diffuse xantholasmique. Ann Dermatol Syphiligr 1881; 47:255–261.
2. Darier J. Pseudoxanthoma elasticum. Monatsschr Prakt Derm 1986; 23:609–614.
3. Lebwohl MG, et al. Classification of pseudoxanthoma elasticum: Report of a consensus conference. J Am Acad Dermatol 1994; 30:103–107.
4. Pope FM. Historical evidence for the genetic heterogeneity of pseudoxanthoma elasticum. Br J Dermatol 1975; 92:493–509.
5. Christiano AM, Lebwohl MG, Boyd CD, et al. Workshop on pseudoxanthoma elasticum: molecular biology and pathology of the elastic fibers. J Invest Dermatol 1992; 99:660–663.
6. Blumcke S, Langness U, Liesegang B, et al. Light and electron microscopic and element analysis of pseudoxanthoma elasticum (Darier-Gronblad-Strandberg Syndrome). Beitr Pathol Bd 1974; 152:179–199.
7. Lebwohl MG, Halperin J, Phelps RG. Brief report: Occult pseudoxanthoma elasticum in patients with premature cardiovascular disease. N Engl J Med 1993; 329:1237–1239.
8. Lebwohl MG, Distefano D, Prioleau PG, Uram M, Yannuzzi LA, Fleischmajer R. Pseudoxanthoma elasticum and mitral-valve prolapse. N Engl J Med 1982; 307: 228–231.
9. Lebwohl MG, Phelps RG, Yannuzzi LA, et al. Diagnosis of pseudoxanthoma elasticum by scar biopsy in patients without characteristic skin lesions. N Engl J Med 1987; 317:347–350.
10. Clarkson JG, Altman RD. Angioid streaks. Surv Ophthalmol 1982; 26:235–246.
11. Pope FM. Autosomal dominant pseudoxanthoma elasticum. J Med Genet 1974; 11:152–157.
12. Danielsen L, Kobayashi T. Pseudoxanthoma elasticum. Acta Dermatovenerol (Stockh) 1974; 54:212–218.
13. Lebwohl MG Schwartz E, Lemlich G, et al. Abnormalities of connective tissue components in lesional and nonlesional tissue of patients with pseudoxanthoma elasticum. Arch Dermatol Res 1993; 285:121–126.
14. Hashimoto K, Klingmuller G, Rodermund O-E. Hyalinosis cutis et mucosae. An electron microscopic study. Acta Derm Venereol 1972; 52:179.
15. Kobayasi T, Wolf-Jurgensen P, Danielsen L. Ultrastructure of shagreen patch. Acta Derm Venereol 1973; 53: 275.
16. Lapiere ChM, Nusgens B, Pierard G, et al. The involvement of procollagen in spatially orientated fibrogenesis. In: Burleigh PMC, Poole AR, eds. Dynamics of connective tissue macromolecules. Amsterdam: North Holland Publishing Company, 1975:33.
17. Martinez-Hernandez A, Huffer WE. Pseudoxanthoma elasticum: Dermal polyanions and the mineralization of elastic fibers. Lab Invest 1974; 31:181–186.
18. Smith JG, Sams W, Davidson EA, Clark RD. Pseudoxanthoma elasticum: Histochemical and biochemical alterations. Arch Dermatol 1962; 86:714–744.
19. Smith JG, Davidson EA, Taylor RW. Cutaneous acid mucopolysaccharides in Pseudoxanthoma elasticum. J Invest Dermatol 1961; 43:429–430.
20. Meyer K, Davidson EA, Linker A, Hoffman P. The acid mucopolysaccharides of connective tissue. Biochim Biophys Acta 1900; 21:506–518.
21. Gordon SG, Overland M, Foley J. Evidence for increased protease activity secreted from cultured fibroblasts from patients with pseudoxanthoma elasticum. Connect Tissue Res 1978; 6:61–68.
22. Kobayasi T, Danielsen L, Asboe-Hansen G. Hyaluronate microfibrils in human dermis. Acta Derm Venereol 1971; 51:27.
23. Lebwohl MG, Longas MO, Konstadt J, et al. Hyuluronic acid and dermatan sulfate in non-lesional pseudoxanthoma elasticum skin. Clin Chim Acta 1900; 238:101–107.
24. Longas MO, Wisch P, Lebwohl MG, Fleischmajer R. Glycosamines of skin and urine in pseudoxanthoma elasticum: evidence for chondroitin-6-phosphate alteration. Clin Chim Acta 1986; 155:227–236.
25. Lebwohl MG, Schwartz E, Feinberg E, Phelps R. Effects of exogenous calcium and glycosaminoglycans on elastic tissue. J Invest Dermatol 1993; 100:530.
26. Uitto J, Lindy S, Turto H, Danielsen L. Biochemical characterization of pseudoxanthoma elasticum: Collagen biosynthesis in the skin. J Invest Dermatol 1971; 87: 44–48.
27. Lebwohl MG, Schwartz E, Lemlich G, et al. Abnormalities of connective tissue components in lesional and nonlesional tissue of patients with pseudoxanthoma elasticum. Arch Dermatol Res 1993; 285:121–126.
28. Cleary EG. The microfibrillar component of the elastic fibers. Morphology and biochemistry. In: Uitto J, Perejda AJ, eds. Connective tissue disease. Molecular pathology of the extracellular matrix. 55–81.
29. Serafini-Fracassini A. Elastogenesis in embryonic and post-natal development. In: Ruggeri A, Motta PM, eds. Ultrastructure of the connective tissue matrix. The Hague: Nijhoff, 140–150.
30. Keene DR, Maddox BK, Kuo HJ, et al. Extraction of extendable beaded structures and their identification as fibrillin-containing extracellular matrix microfibrils. J Histochem Cytochem 1991; 39:441.
31. McKusick V. Heritable disorders of connective tissue, 4th ed. St Louis: CV Mosby, 1972.

32. Pyeritz RE, McKusick V. The Marfan Syndrome: Diagnosis and management. N Engl J Med 1979; 300:772–777.
33. Lee B, Godfrey M, Vitale E, et al. Linkage of Marfan syndrome and a phenotypically related disorder to two fibrillin genes. Nature (Lond). 1991; 352:330–334.
34. Dietz HC, et al. Marfan syndrome caused by recurrent de novo missense mutation in the fibrillin gene. Nature (Lond). 1991; 352:337–339.
35. Mariencheck MC, Davis EC, Zhang H, et al. Fibrillin-1 and fibrillin-2 show temporal and tissue-specific regulation of expression in developing elastic tissues. Connect Tissue Res 1995; 31,2:87–97.
36. Cleary EG, Gibson MA. Elastin-associated microfibrils and microfibrillar proteins. Int Rev Connect Tissue Res 1983; 10:97–209.
37. Zhang H, Apfelroth SD, Hu W, et al. Structure and expression of fibrillin-2, a novel microfibrillar component preferentially expressed in elastic matrices. J Cell Biol 1994; 124:855–863.
38. Pereira L, D'Alessio M, Ramirez F, et al. Genomic organization of the sequence coding for fibrillin, the defective gene product in Marfan syndrome. Hum Mol Genet 1993; 2:961–968.
39. Davis CG. The many faces of epidermal growth factor repeats. New Biol 1990; 2:410–419.
40. Glanville RW, Qian RQ, McClure DW, Maslen CL. Calcium binding, hydroxylation, and glycosylation of the precursor epidermal growth factor-like domains of fibrillin-1, the Marfan gene protein. J Biol Chem 1994; 269:26630–26643.
41. Corson GM, Charlberg SC, Dietz HC, et al. Fibrillin binds calcium and is code by cDNA that reveals multidomain structure and alternatively spliced exons at the 5′ end. Genomics 1993; 17:476–484.
42. Godfrey M, Olson S, Burgio RG, et al. Unilateral microfibrillar abnormalities in a case of asymmetric Marfan syndrome. Am J Genet 1990; 46:661–671.
43. Hollister DW, Godfrey M, Sakai LY, Pyeritz RE. Immunohistologic abnormalities of the microfibrillar-fiber system in the Marfan syndrome. N Engl J Med 1990; 323:152–159.
44. Aoyama T, Franke U, Dietz HC, Furthmayr H. Quantitative differences in biosynthesis and extracellular deposition of fibrillin in cultured fibroblast distinguish five groups of Marfan syndrome patients and suggest distinct pathogenetic mechanisms. J Clin Invest 1994; 94:130–137.
45. Milewicz DM, Pyeritz RE, Crawford ES, Byers PH. Marfan syndrome: Defective synthesis, secretion and extracellular matrix formation of fibrillin by cultured dermal fibroblast. J Clin Invest 1992; 89:79–86.
46. Beals RK, Hecht F. Congenital contractural arachnodactyly. A heritable disorder of connective tissue. J Bone Joint Surg 1971; 53A,5:987–993.
47. Tsipouras P, Del Mastro R, Sarfarazi M, et al. Genetic linkage of Marfan syndrome, ectopia lentis and congenital contractural arachnodactylyl to the fibrillin genes on chromosome 15 and 5. N Engl J Med 1992; 326:905–909.
48. Putnam EA, Zhang H, Ramirez F, Milewicz DM. Fibrillin-2 mutations in the Marfan-like disorder, Congenital Contractural Arachnodactyly. Nature Genet, in press.
49. Raybold MC, Birley AJ, Moss C, et al. Exclusion of an elastin gene (ELN) mutation as the cause of pseudoxanthoma elasticum (Pseudoxanthoma Elasticum) in one family. Clin Genet 1994; 45:48–51.
50. Godfrey M, Cisler J, Geerts M-L, et al. Fibrillin immunofluorescence in pseudoxanthoma elasticum. J Am Acad Dermatol 1995; 32:589–594.
51. Fleischmajer R, Contard P, Schwartz E, et al. Elastin-associated microfibrils (10 nm) in a three-dimensional fibroblast culture. J Invest Dermatol 1991; 97:638–643.
52. Fleischmajer R, MacDonald ED, Contard P, Perlish JS. Immunohistochemistry of a keratinocyte-fibroblast coculture model for reconstruction of human skin. J Histochem Cytochem 1993; 41,9:1359–1366.
53. Sapadin A, Schwartz E, Lebwohl MG. Fibrillin 5 synthesis in cultured pseudoxanthoma elasticum fibroblasts. J Invest Dermatol 1994; 102:575.

Meta-analysis:

An Update

HENRY S. SACKS, PH.D., M.D., DINAH REITMAN, M.P.S., DANIEL PAGANO, B.A., AND BRUCE KUPELNICK, B.A.

Abstract

A fairly new type of research, termed meta-analysis, attempts to analyze and combine the results of previous reports. In 1992 we updated our 1987 survey of 86 meta-analyses of randomized control trial reports in the English language literature with an additional 78. We evaluated the quality of these meta-analyses using a scoring method that lists 23 items in six major areas: study design, combinability, control of bias, statistical analysis, sensitivity analysis, and application of results. Of the 23 individual items, the mean number satisfactorily addressed was 7.63 ± 2.84 (mean ± S.D.) for 40 papers published from 1955 through 1982, 6.80 ± 3.86 for 66 papers published from 1983 through 1986, and 11.91 ± 4.79 for 58 papers published from 1987 through 1990 ($F = 31.3$, $p < .001$). We noted that methodology has definitely improved since our first survey of meta-analyses, but an urgent need still exists for a better search of the literature, quality evaluation of trials, and a synthesis of the results. Recently, meta-analysis has expanded to cover non-randomized studies, including evaluation of diagnostic tests and pooling of epidemiologic studies. There is growing concern for standards, and several methodologic issues remain unresolved.

IN AN ARTICLE IN THE *NEW YORK TIMES*, L. K. ALTMAN (1) stated:

> A meta-analysis aims at gleaning more information from existing data by pooling the results of many smaller studies and applying one or more statistical techniques. The benefits or hazards that might not be detected in small studies can be found in a meta-analysis that uses data from thousands of patients.

Meta-analysis is nearing its centennial: the earliest meta-analysis of which we are aware, published in 1904, combined the results of five studies on the effectiveness of inoculation on enteric fever (2). The next meta-analysis that we found was a paper published in 1955 (3) by Henry K. Beecher on the powerful effects of the placebo, which is the source of the frequently cited estimate that placebos are effective one third of the time, no matter what the problems they are used for.

The number of published meta-analyses has burgeoned in the past 20 years or so. Since the publication of our last survey (4), there have been 572 citations in Medline of "meta-analysis" and "randomized controlled trials(s)" or "controlled trial(s)," 156 from 1990 through 1992 and 416 from 1993 through 1996. A file is being maintained by our collaborative group at Metaworks. As of March 1996, it consists of 1237 meta-analyses of randomized control trials (RCTs), 89 meta-analyses of non-RCTs, and 590 epidemiologic and 155 diagnostic meta-analyses.

In our 1987 report on meta-analyses of RCTs, we listed the purposes of meta-analyses, including the following:

1. to increase statistical power for primary endpoints and for subgroups,

---

This article is dedicated to the memory of Dr. Thomas C. Chalmers, president and dean of Mount Sinai from 1973 to 1983, and a major contributor to many areas of medical research, including meta-analysis.

From the Thomas C. Chalmers Clinical Trials Unit and Departments of Community Medicine, Medicine, and Biomathematical Sciences, Mount Sinai School of Medicine of CUNY, New York, NY and Metaworks, Inc., Boston, MA. Address requests for reprints to Dr. Henry S. Sacks, Thomas C. Chalmers Clinical Trials Unit, The Mount Sinai Medical Center, 1 Gustave L. Levy Place, New York, NY 10029.

2. to resolve uncertainty when reports disagree,
3. to improve estimates of effect size,
4. to answer questions not posed at the start of individual trials, and
5. to bring about improvements in the quality of the primary research (5).

Although at that time meta-analyses in health care had been applied largely to clinical trials, they had not been confined solely to this area. Meta-analyses had been applied to epidemiologic research as well as to diagnostic tests; these areas have grown recently.

Survey of Meta-analyses of Randomized Control Trials

In 1987 (5) and again in 1992 (4) surveys of meta-analyses of RCTs were conducted by this research group and evaluated using a quality scoring system consisting of six major areas: study design, combinability, control of bias, statistical analysis, sensitivity analysis, and problems of applicability.

For these surveys, the English language medical literature (from January 1966 through December 1990) was searched for papers that pooled the results of controlled clinical trials. Papers were found in Current Contents and by computer searches of the National Library of Medicine (NLM) and Bibliographic Retrieval Services, Inc. (BRS) databases by looking for reviews of specific subjects and using the terms "meta-analysis," "pooled or pooling," or "combined or combining" in title, abstract or, where available, full text. Other sources of papers included references in papers found by these methods and by personal communication.

Criteria for inclusion of papers in our analyses were that data from more than one study must be combined, and at least one of the studies pooled must be a randomized clinical trial. Each paper was evaluated independently by two investigators using a scoring sheet which lists what we consider to be the important elements of a meta-analysis. The evaluators were blinded as to the name of the journal and the authors of the papers.

From a total of 177 papers retrieved, 86 studies met the inclusion criteria in our first survey. Subsequently, we found an additional 78. When the same results were reported or updated in more than one publication, only the latest publication was cited, but the evaluation was based on a composite of all available information.

We found that, after an initial learning period, two evaluators agreed on more than 90% of items scored; most disagreements were caused by oversights.

The scoring results are listed in **Table 1** and a discussion of each major area follows. We divided the papers into three time periods, selected to give roughly equal numbers in each: 1955 through 1982, 1983 through 1986, and 1987 through 1990.

Study Design

In meta-analysis, like any other form of research, it is important to try to make the process as rigorous and well defined as possible. Study design consists of preparation of a protocol, literature search, list of trials analyzed and log of rejected trials, treatment assignment, ranges of patients, diagnoses, and treatments.

Protocol

Preparation of a protocol refers to establishing a research plan beforehand which consists of defining the questions to be answered, the inclusion criteria of the study and the methodology to be used.

Only a minority of the meta-analyses in both our surveys clearly indicated that they devised a protocol beforehand.

Literature Search

Details of a literature search should be provided which consist of more than one method, i.e., a computer search in addition to at least one other search procedure, e.g., Current Contents, references in review articles or in other clinical trials to be included in the analysis, or consulting experts in the field. It has been shown that relying solely on computer searches will only capture 66% of the available studies (6, 7). Only 35% of the papers reviewed in the first survey as opposed to 69% from 1987–1990 conducted exhaustive literature searches.

List of Trials Analyzed and Log of Rejected Trials

This is fairly straightforward, i.e., the meta-analyst should provide a list of the trials that were included in the analysis as well as those that were excluded and the reasons why. Although most meta-analyses provide a list of primary studies that were included, even in the later time period, less than half provided a rejection log.

TABLE 1

Comparison of Quality Features among Meta-analyses

	1955—1982 (n = 40) Adequate	1983–1986 (n = 66) Adequate	1987–1990 (n = 58) Adequate
	Number (%)		
Prospective design			
Protocol	5 (13)	2 (3)	13 (22)
Literature search	10 (25)	24 (36)	40 (69)
List of trials analyzed	37 (93)	56 (85)	54 (93)
Log of rejected trials	4 (10)	11 (17)	24 (41)
Treatment assignment	38 (95)	17 (26)	46 (79)
Ranges of patients	13 (33)	12 (18)	36 (62)
Ranges of treatment	20 (50)	25 (38)	39 (67)
Ranges of diagnosis	18 (45)	21 (32)	34 (59)
Combinability			
Criteria	17 (43)	26 (39)	39 (67)
Measurement	5 (13)	17 (26)	27 (47)
Control of bias			
Selection bias	0	0	7 (12)
Data-extraction bias	0	0	7 (12)
Interobserver bias	0	7 (11)	11 (19)
Source of support	17 (43)	14 (21)	16 (28)
Statistical analysis			
Statistical method	22 (55)	40 (61)	45 (78)
Statistical errors	15 (38)	31 (47)	38 (66)
Confidence intervals	14 (35)	27 (41)	49 (84)
Subgroup analysis	28 (70)	39 (59)	45 (78)
Sensitivity analysis			
Quality assessment	9 (23)	10 (15)	15 (26)
Varying methods	6 (15)	12 (18)	25 (43)
Publication bias	3 (8)	11 (17)	24 (41)
Application of results			
Caveats	37 (93)	31 (47)	41 (71)
Economic impact	0	2 (3)	3 (5)

Treatment Assignment

The most important question bearing on the validity of the data pooled is the method of treatment assignment in the primary studies. It has been shown that results of trials using historical controls are more apt to favor the new treatment than results of the same therapy tested in RCTs (8). The proportion of papers using data only from RCTs has declined.

Ranges of Patients, Treatment, and Diagnosis

For the reader to judge the validity and generalizability of a meta-analysis, data should be provided on the patients, diagnoses, therapies, and endpoints in the original studies. The ranges of patient characteristics in all the trials analyzed—age, sex, relevant socioeconomic data, other diseases, etc.—should be included. Details on these items were given in only a minority of the meta-analyses in the first two time periods. The ranges of treatments should also be defined. Did all patients in all trials receive the same, or similar, therapy? In meta-analyses of drug therapies, were trials combined which used the same drug, or the same class of drug? Dosage and route, as well as frequency and duration of therapy in all the trials to be pooled, should be provided. This criterion was variably fulfilled in the studies analyzed. Similarly, data should be presented on the ranges of diagnoses in the pooled trials. Were diagnostic criteria the same in all trials? What stages or grades of disease were included? Such details on this important question were reported in less than half the meta-analyses in the first two time periods, but in greater than two thirds in the third time period.

Combinability

A major issue in pooling data is whether the results of the separate trials can be meaningfully combined; this should be explicitly addressed by the meta-analyst, in sufficient detail for the reader to determine that a useful and clinically relevant result will be obtained. The items under this heading are criteria for pooling and measurement.

Criteria

What criteria were used to decide that the studies analyzed were similar enough to be pooled? The meta-analyst should note any differences in the primary studies, and discuss how these differences affect the conclusions of the meta-analysis. Less than half of the meta-analysts in the first two time periods and 59% in the third detailed their criteria for pooling.

Measurement

Related to the problem of combinability is the statistical issue of heterogeneity. In addressing questions of combining estimates from different studies, statisticians distinguish between two possible models. In model 1 (9), "fixed effects", each study is considered to be a sample from the same population and provides an estimate of a single, underlying true rate, with the differences due to experimental error (within study variability). In model 2 (10), "random effects" each study is considered to be from a different population, the rate varies from study to study, and their differences are due to experimental error and to differences in the populations (between study variability). There are a variety of methods for deciding which of the models above is more appropriate (11). In either case, there are tests for homogeneity which help decide the degree of caution with which pooled results should be interpreted. Evidence of a statistical test for homogeneity was found in few of the meta-analyses in the first time periods, but rose to nearly half in the most recent period.

Control and Measurement of Potential Bias

When performing a meta-analysis, as in any scientific endeavor, potential sources of unconscious bias should be controlled where possible. The sources of bias to consider are selection, data collection, interobserver, and source of support.

Selection Bias

To avoid bias in the selection and rejection of papers, the decision to include a paper should be made by looking only at its methods and not at its results, or by looking at the two separately under coded conditions. This important source of bias was not reported in any of the meta-analyses reviewed for our first survey and in only 7% in the later period.

Data Extraction Bias

As with any other data-gathering process that requires interpretation, observers may disagree. When papers list a variety of subgroups, endpoints, exclusions, etc., it is quite possible that readers may vary in how they extract the data from a particular study. The ideal way to control for this type of bias is to have the data extracted by more than one observer, each of whom is blinded to the various treatment groups through a coded photocopying process, and then measure the *interobserver agreement*. In none of the meta-analyses in the first period was such agreement reported, but interobserver agreement seems to be increasing recently. The data extraction process was not blinded in any of the earlier papers, and in only 12% of those in the latest period.

Source of Support

It is useful for the reader to know who financed a study when deciding how much credence to give to its conclusions. Potential conflicts of interest do not necessarily disqualify a study, but they should be clearly acknowledged. The source of support was specified in a minority of papers and, if anything, appears to have declined over time.

Statistical Analysis

This category deals with questions of statistical methodology. The items that are relevant are methods, statistical errors, confidence intervals, and subgroup analyses.

Methods

We evaluated as "adequate" any recognized method of pooling, except simple addition of successes across all trials to give an overall average which was rated as "partial." An adequate method was used in more than half of the meta-analyses; this rate seems to have increased over time.

The most commonly used method was the Mantel-Haenszel (12) test or a modification thereof (13). Other studies combined data by calculating a standardized average effect size; many also performed various types of regression analyses and/or significance tests. A few papers used various other methods of pooling, and five papers did not specify the methods used. For further discussion of these various methods, see "Remaining Problems."

Statistical Errors

Less than half of the meta-analyses in the first two periods showed an awareness of the potential problems of the Type I (concluding that there is a difference when none exists) and Type II (concluding that there is no difference when there is one) statistical errors. There has been some improvement in the last 3 years.

Confidence Intervals

It is often more useful to the reader to have an estimate with confidence intervals of the difference between the success rates of the modalities being compared, than to have only the results of significance tests. Confidence intervals for major outcomes were given in less than half (43%) of the meta-analyses in the first periods, but this has risen to 84% recently.

Subgroup Analyses

One of the purposes of meta-analysis, as previously stated, is to increase the statistical power for subgroup analyses. Relevant subgroups were analyzed in most papers.

Sensitivity Analysis

Depending on the test chosen, the same set of data can be combined to give different conclusions (14). Similarly, the results may vary depending on the overall quality of the primary trials and on whether certain trials, subgroups of patients, or other important variables are excluded or changed.

Quality Assessment

In a meta-analysis, the methodologic rigor and scientific quality of the papers to be combined should be assessed and considered in formulating recommendations (15, 16). If the original methodology is poor, the resulting conclusion is less reliable. Such features as the randomization process, the measurement of patient compliance, the blinding of patients and observers, the statistical analyses, and the handling of withdrawals in each primary study should be examined. This issue of quality was fully addressed in only a small proportion of the meta-analyses and was not even mentioned in almost half. There was no evidence of improvement over time.

Varying Methods

Each meta-analysis should include in a sensitivity analysis data which show how the results vary through the use of different assumptions, tests, and criteria. This type of analysis was performed in only 15% of the meta-analyses reviewed in the first period, but the percentage increased appreciably in the second and third.

Publication Bias

One of the criticisms sometimes made of meta-analysis is that there may be some unpublished studies that would contradict the results of published studies, and there is some evidence that negative studies are less likely to be published than positive ones (17, 18). A simple method has been proposed for calculating the number of unpublished negative studies required to refute the published evidence (19), which is one possible measure of the strength of the published evidence. In only 8% of meta-analyses was publication bias considered in the first period, compared with 17% in the second and 41% in the third period.

Application of Results

Caveats

Once the results of the pooling process are available, the meta-analyst should attempt to put them into perspective, based on all of the considerations above. Does the new therapy seem to be established as more effective than the old for all patients, for some subgroups, etc.? Or should the conclusions be taken only as suggestions for future study? Such caveats were included in the discussion in varying proportions of the papers with, if anything, a decline over time.

Economic Impact

In today's climate of financial constraints on health care expenditure, it is increasingly important to consider the economic impact of adopting new modalities of treatment or diagnosis. Although some may consider this a topic for other studies, we were disappointed that less than 5% of the meta-analyses included a thorough analysis of economic impact.

Of the 23 individual items, 7.63 ± 2.84 (mean ± S.D.) were addressed adequately in the 40 meta-analyses published between 1955 and 1982, 6.80 ± 3.86 were addressed adequately in the 66 published between 1983 and 1986, and 11.91 ± 4.79 were addressed adequately in the 58 published between 1987 and 1990 ($F = 37.3$ $p < .001$). Twenty-two of the 58 papers published in the last period referred to our initial survey or to other similar guides for meta-analysis (which suggested that the guidelines were being used).

Remaining Problems

The greatest problem is defining the role of meta-analysis. When should it be attempted and how should its results be used? We believe that the best way to answer questions about the efficacy of new therapeutic or diagnostic modalities is to perform well-designed, adequately sized randomized control trials. Meta-analysis may have a role when definitive trials are impossible or impractical, when trials have been performed but the results are inconclusive or conflicting, or while awaiting the results of definitive studies. Meta-analysis, like decision analysis, can give quantitative estimates of the weight of available evidence that can be helpful in making clinical decisions. There is, however, a danger that meta-analysis may be used inappropriately or indiscriminately. As with many other types of analysis, the quality of the results depends on the quality of the input. Therefore, the question posed in each meta-analysis should be explicitly stated and clinically relevant.

Difficulty still exists in locating both meta-analyses and RCTs in the literature because present literature-searching and indexing systems do not always distinguish primary studies and reviews from meta-analyses, or RCTs from other clinical trials. Thus it cannot be claimed that the papers found for these analyses are an exhaustive or representative sample. It is also quite likely that there are unpublished meta-analyses. Investigators will facilitate the process of recovery and integration of important clinical information if they insist on inclusion of the terms, "randomized" and "meta-analysis" in titles and abstracts so that indexing can be improved.

More attention should also be paid to statistical issues and the advantages and disadvantages of the various pooling methods. A variety of statistical techniques have been developed for combining the results of separate studies. For example, there are several methods for combining the probability values or test statistics from individual studies (14). These methods, however, may not distinguish between small studies with large effects and large studies with small effects, and they do not yield an estimate of the size of the effect. The Mantel-Haenszel method (12) or a modification of that technique for combining separate two-by-two tables (Yusuf-Peto method) (13) is becoming increasingly popular. These methods have several useful properties: they compare each treatment only with its own control and weight studies according to their sample size. They can include a test for homogeneity, as well as an estimate of the effect size, but they may have some undesirable properties (see below). In the psychiatric literature, studies have been combined by computing effect sizes, defined as the mean difference between experimental and control groups divided by the control group standard deviation (20). This method allows for the pooling of different endpoints, because all findings are transformed into common units, but the conclusions may be difficult to interpret clinically; the validity of this process is, thus, open to question. A few papers have used multivariate methods or log-linear models to attempt to adjust for differences between studies.

Since the publication of our first paper, there have been several important contributions to the statistical methods in meta-analysis. The validity of the Yusuf-Peto one-step method has been challenged (21). Simulations have confirmed this (22), and in those experiments the most valid method of determining variances has been described by Robins et al. (23). When zero observations in one or both groups present a problem, the exact method, as modified and automated by Mehta (24) is optimal. A Bayesian approach has been advocated (25). However, we encountered no published meta-analyses that used the last two methods.

Another important development in meta-analysis was a textbook entitled "Effective Care in Pregnancy and Childbirth" (26) with more than 1,000 individual meta-analyses. These have not been included in our survey because they represent a special case—these studies were written on consignment by the editors and based on a database of RCTs that had been collected in Oxford (27). Obviously this approach needs to be replicated by the conduct of large numbers of meta-analyses in other fields of medicine.

Some meta-analysts believe that the only valid pooling of results should include the outcomes for all randomized patients regardless of how long (or even whether) they received the assigned treatment (the intention-to-treat method) (13). The authors of at least one paper apparently felt that dropouts and withdrawals should be excluded (exclusion method) (28). However, the data on withdrawals, dropouts, etc. are not always available. We believe that the results should be reported both according to the intention to treat and exclusion rules to facilitate evaluation of the differences. If there are none, there is no problem. If clinically important differences exist the study may be difficult to interpret (29).

Due to the problem of publication bias, some meta-analysts choose to supplement their published data with unpublished trials and/or data. Unpublished results may be less reliable, because they have not been found acceptable by peer-reviewers and may not be collected with the same rigor or accuracy as published results (30). However, the potential problems inherent in unpublished results make it unclear as to whether both types of data should be given equal weight.

Greater uniformity in reporting meta-analyses would be helpful to readers. Many of the meta-analyses found were written in the standard format of scientific papers with detailed methods and results sections, but several were editorials, leading articles, or letters to the editor with little detail on methodology. Meta-analyses should be presented with sufficient information for readers to draw their own conclusions about the validity of the results.

With the growth in the number of published meta-analyses, there has been growing concern about their quality (31). A number of national and international conferences have been held to develop and refine standards for meta-analysis (32, 33). The Potsdam International consultation on meta-analysis brought together proponents and critics of meta-analysis for a lively discussion which was published in 1995 (33). It concluded with methodologic guidelines which should be consulted by anyone wishing to perform a meta-analysis. We quote here their guiding principles, which we endorse (34).

1. A systematic review must address a specific health care question. The question will determine which studies and data are relevant, and how they should be synthesized.
2. Methodology must serve biology and the users and providers of health care. Therefore, a team with expertise in both the content area and methodology is ideally suited to conduct valid, useful systematic reviews.
3. A systematic review requires collaboration with the investigators who conducted the primary studies.
4. Systematic reviews are retrospective research and are potentially subject to many of the same biases that affect other retrospective studies; therefore, a systematic review has to rely on both good randomized control trial methodology and good review methodology.
5. For several reasons, review methods may vary (e.g., scarce resources may limit search strategies). Thus, the review methods actually used must be described in detail.
6. The existence of unsatisfactory randomized trials, case-control studies, and cohort studies does not mean that any of these study designs should be abandoned; it means that they should be critically appraised, empirically studied, and improved. Overviews of observational studies require a great deal of methodologic development.

Validity

A major unresolved issue is the validity of meta-analysis. Ingram Olkin, a statistician who has developed methodologies for meta-analysis, has proposed the following ranking of the validity of evidence from various study designs (35):

Hierarchy of Evidence
1. Anecdotal case reports
2. Case series without controls
3. Series with literature controls
4. Analyses using computer data bases
5. Case control observational studies
6. Series based on historical control groups
7. Simple randomized control trials
8. Confirmed randomized control trials including META-ANALYSIS
9. Meta-analysis with original data

Others are much less optimistic. Alvan Feinstein complained that meta-analyses frequently violate scientific principles of precision, homogeneity, and consistency, while focusing on the big process of aggregation, but not on the overall small processes that produced the primary data (36).

We agree that a meta-analysis must be carefully done, according to the criteria discussed above, and that careful assessment of the primary data is a critical step. However, we believe meta-analysis can provide important and useful information on which to base clinical decisions. No question can ever be considered completely closed, and thoughtful physicians will always be searching for new data and revising their strategies.

Meta-analysis of Diagnostic Tests

Irwig et al. (37) proposed methodologic standards for meta-analyses of diagnostic tests. They suggested that the process should include six steps:

1. Determine the objective and size of the meta-analysis. This should include a clear statement of the test of interest, the disease of interest, and the reference standard by which it is measured, the clinical question and context, i.e., is the objective to evaluate a single test or to compare the accuracy of different tests?
2. Retrieve the relevant literature. The literature retrieval procedure should be described with search and link terms given and inclusion and exclusion criteria stated.
3. Extract and display the data. Studies should be assessed by two or more readers. The authors should explain how disagreements between readers were resolved. The meta-analysis should give a full listing of diagnostic accuracy and study characteristic for each primary study?
4. Estimate diagnostic accuracy. The method of pooling sensitivity and specificity should take account of their independence. When multiple test categories are available, they should be used in the summary.
5. Assess the effect of variation in study validity on estimates of diagnostic accuracy. The relationship between estimates of diagnostic accuracy and study validity of the primary studies should be examined for each of the following design characteristics: appropriate reference standard, independent assessment of the test or tests, and reference standard avoidance of verification bias. In comparative studies, were either all of the tests of interest applied to each patient or were patients randomly allocated to the tests? Are analytic methods used that estimated whether study design flaws affect diagnostic accuracy rather than just test threshold?
6. Assess the effect of variation in the characteristics of patients and test on estimates of diag-

nostic accuracy (generalizability). Is the relation examined between estimates of diagnostic accuracy and characteristics of the patients and test? Are analytic methods used which differentiate whether characteristics affect diagnostic accuracy or test threshold?

Irwig et al. (37) applied these standards to eleven published meta-analyses and found deficits both in the meta-analyses and in the primary studies on which they were based.

Meta-analyses of Observational Studies

This is an area of much debate, where again a critical factor is the methodologic quality of the primary studies. Meta-analyses of observational studies have been done in many areas and have produced intriguing, but controversial findings, including that chlorination of water supplies may slightly, but significantly, increase the incidence of malignancies of several organs (38); that increased fat in the diet does not increase the incidence of breast cancer (39). It is clear that the danger of incorrect conclusions is higher than when combining results of RCTs, but a distinguished group of researchers concluded that meta-analysis must not be limited to RCTs (40). Non-RCTs data make up the bulk of what is known regarding medical care. It is imperative that technology assessors, policy makers, clinicians, and patients have clear and explicit means for summarizing all sorts of published data. Policy makers cannot afford to ignore the vast majority of information available today merely because it does not fit the simple model of classical meta-analysis. Researchers must develop new methods and new models of analysis to assist those who must use today's imperfect data to make difficult decisions today.

Meta-analysis Software

At least three computer software packages for meta-analysis are available. A detailed review and comparison was recently published (41).

In conclusion, if meta-analysis is to be accepted as a scientific tool, each meta-analysis should be conducted like a scientific experiment, beginning with a clear plan of the question to be answered and the methodology to be used. Attention must be paid to intraobserver and interobserver variability, and attempts should be made to identify and minimize bias. Concern has been expressed about the validity of pooling (42, 43) but the process is used increasingly and defended frequently (44–47). Quantitative synthesis of the data in similar RCTs is more useful to the practicing physician than a traditional narrative review article, but such syntheses must be performed properly to warrant serious attention. We hope that the points we have raised here will stimulate discussion that will ultimately lead to better meta-analyses. Once the method has matured, we believe it will be considered as superior to classical review articles, as the randomized control is to the historical control in therapeutic trials.

References

1. Altman LK. Study shows wider utility for aspirin. New York Times 1994 Jan 7:A20.
2. Pearson K. Report on certain enteric fever inoculation statistics. Br Med J 1904; 3:1243–1246.
3. Beecher HK. The powerful placebo. JAMA 1955; 159: 1602–1606.
4. Sacks HS, Berrier J, Reitman D, Pagano D, Chalmers TC. Meta-analysis of randomized control trials: An update. In: Balder WC 3rd, Mosteller F, eds. Medical uses of statistics, 2nd ed. Boston, MA: NEJM Books, 1992:427–442.
5. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. N Engl J Med 1987; 316:450–455.
6. Poynard T, Conn HO. The retrieval of randomized clinical trials in liver disease from the medical literature. A comparison of MEDLARS and manual methods. Controlled Clin Trials 1985; 6:271–279.
7. Dickersin K, Hewitt P, Mutch L, Chalmers I, Chalmers TC. Perusing the literature: Comparison of MEDLINE searching with a perinatal trials database. Controlled Clin Trials 1985; 6:306–317.
8. Sacks H, Chalmers TC, Smith H, Jr. Randomized versus historical controls for clinical trials. Am J Med 1982; 72:233–240.
9. Laird N, Mosteller F. Some statistical methods for combining experimental results. Int J Technol Assess Health Care 1990; 6:5–30.
10. DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled Clin Trials 1986; 7:177–188.
11. Berlin JA, Laird NM, Sacks HS, Chalmers TC. A comparison of statistical methods for combining event rates from clinical trials. Stat Med 1989; 8:141–151.
12. Mantel N, Haentzel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959; 22:719–748.
13. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: An overview of the randomized trials. Prog Cardiovasc Dis 1985; 27:335–371.
14. Rosenthal R. Combining results of independent studies. Psychol Bull 1978; 85:185–193.
15. Chalmers TC, Smith H Jr., Blackburn B, et al. A method for assessing the quality of a randomized control trial. Controlled Clin Trials 1981; 2:31–49.
16. DerSimonian R, Charette LJ, McPeek B, Mosteller F. Reporting on methods in clinical trials. N Engl J Med 1982; 306:1332–1337.
17. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H

Jr. Publication bias and clinical trials. Controlled Clin Trials 1987; 8:343–353.

18. Simes RJ. Publication bias: The case for an international registry of clinical trials. J Clin Oncol 1986; 4:1529–1541.
19. Rosenthal R. The "file drawer problem" and tolerance for null results. Psychol Bull 1979; 86:638–641.
20. Glass GV, McGraw B, Smith ML. Meta-analysis in social research. Beverly Hills, CA: Sage, 1981:21–56.
21. Greenland S. Quantitative methods in the review of epidemiologic literature. Epidemiol Rev 1987; 9:1–30.
22. Emerson JD. Confidence intervals for the odds ratio in meta-analysis: The state of the art. Stat Methods Med Res 1994; 3:157–178.
23. Robins J, Greenland S, Breslow NE. A general estimator for the variance of the Mantel-Haenszel odds ratio. Am J Epidemiol 1986; 124:719–723.
24. Mehta CR, Patel NR, Gray R. Computing an exact confidence interval for the common odds ratio in several 2 × 2 contingency tables. J Am Stat Assoc 1985; 80:969–973.
25. Eddy DM, Hasselblad V, Shachter R. A Bayesian method for synthesizing evidence: The confidence profile method. Int J Technol Assess Health Care 1990; 6:31–56.
26. Chalmers I (ed.), Enkin M (ed.), Keirse MJNC. Effective care in pregnancy and childbirth. Oxford: Oxford University Press, 1989.
27. Chalmers I, Hetherington J, Newdick M, et al. The Oxford Database of Perinatal Trials: developing a register of published reports of controlled trials. Controlled Clin Trials 1986; 7:306–324.
28. Levitt SH, McHugh RB, Song CW. Radiotherapy in the postoperative treatment of operable cancer of the breast. Part II. A re-examination of Stjernsward's application of the Mantel-Haneszel statistical method. Evaluation of the effect of the radiation on immune response and suggestions for postoperative radiotherapy. Cancer 1977; 39(Suppl):933–940.
29. Bhaskar R, Reitman D, Sacks HS, Smith H Jr., Chalmers TC. Loss of patients in clinical trials that measure long-term survival following myocardial infarction. Controlled Clin Trials 1986; 7:134–148.
30. Relman AS. News reports of medical meetings: how Reliable are abstracts? [editorial]. N Engl J Med 1980; 303: 277–278.
31. Moher D, Olkin I. Meta-analysis of randomized controlled trials. A concern for standards. JAMA 1995; 274:1962–1964.
32. Boissel JP, Sacks HS, Leizorovicz A, Blanchard J, Panak E, Peyrieux JC. Meta-analysis of clinical trials: Summary of an international conference. Eur J Clin Pharmacol 1988; 34:535–538.
33. Potsdam International Consultation on Meta-analysis. J Clin Epidemiol 1995; 48:1–169.
34. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. J Clin Epidemiol 1995; 48:167–171.
35. Olkin I. Statistical and theoretical considerations in meta-analysis. J Clin Epidemiol 1995; 48:133–46.
36. Feinstein AR. Meta-analysis: Statistical alchemy for the 21st century. J Clin Epidemiol 1995; 48:71–79.
37. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med 1994; 120:667–676.
38. Morris RD, Audet AM, Angelillo IF, Chalmers TC, Mosteller F. Chlorination, chlorination by-products, and cancer: a meta-analysis. Am J Public Health 1992; 82: 955–963.
39. Hunter DJ, Spiegelman D, Adami HO, et al. Cohort studies of fat intake and the risk of breast cancer—a pooled analysis. N Engl J Med 1996; 334:356–361.
40. Haselblad V, Mosteller F, Littenberg B, et al. A survey of current problems in meta-analysis. Discussion from the Agency for Health Care Policy and Research inter-PORT Work Group on Literature Review/Meta-Analysis Med Care 1995; 33:202–220.
41. Normand S-LT. Meta-analysis of software. A comparative review. Am Stat 1995; 49:298–309.
42. Elashoff JD. Combining results of clinical trials [editorial]. Gastroenterology 1978; 75:1170–1172.
43. Goldman L, Feinstein AR. Anticoagulants and myocardial infarction. The problems of pooling, drowning, and floating. Ann Intern Med 1979; 90:92–94.
44. Gerbarg ZB, Horwitz RI. Resolving conflicting clinical trials: Guidelines for meta-analysis. J Clin Epidemiol 1988; 41:503–509.
45. Wachter KW. Disturbed by meta-analysis? Science 1988; 241:1407–1408.
46. L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. Ann Intern Med 1987; 107:224–233.
47. Thacker SB. Meta-analysis. A quantitative approach to research integration. JAMA 1988; 259:1685–1689.

Hemophilia:

Yesterday, Today, and Tomorrow

Louis M. Aledort, M.D.

Abstract

Hemophilia is an ancient disease which was not defined until the 1940s. Marked advances in management have altered the lifestyle of these patients and extended their life. The use of human-derived blood products had brought devastating viral infections to hemophiliacs. New products and creative interventions offer new patients safer and longer life, prophylaxis, continuous infusion, and gene therapy offer a bright future. Inhibitor induction and its management remain a significant challenge, but financial underwriting of care for this chronic disease remains the major uncertainty regarding fulfillment of the scientific and clinical gains.

Although hemophilia as a laboratory diagnosis was not recognized until the 1940s, it was clinically described in ancient Egypt and in the Talmud. Historians have dramatized this entity by detailing the large family of Queen Victoria, and by writing about the impact this disease had on the royal house of Russia. Early treatment included hypnosis, isolation, ice packs, and a very early death from spontaneous bleeding.

Concomitant with the discovery of blood-banking technology was the recognition that hemophiliacs were missing a factor that should be present in plasma. This led to the use of fresh-frozen plasma as the first replacement product to treat bleeding. Its limitation was fluid overload, and thus elective procedures were not easily undertaken. Cryoprecipitate, discovered by Dr. Judith Poole in the 1960s, and now a human plasma derivative, the first concentrated factor VIII product have revolutionized treatment for these patients. Elective surgery and the use of this material at home made substantial beneficial changes in the lives of patients with hemophilia. At the same time, comprehensive care, as an approach to the management of this chronic lifelong disease, became the accepted approach to the care of patients with hemophilia (1). Self or family infusion proliferated, and patients became emancipated from years of being treated in emergency rooms or in physicians' offices. Full-time education, a brighter outlook for productivity, job satisfaction, and a long life became possible. This multidiscipline approach has become the model for chronic disease care.

By the early 1970s, the major concerns of patients and physicians were access to care, appropriate diagnosis, home care training, and whether treatment and blood product utilization would be underwritten by the health care system. By the late 1970s, blood fractionation technology had advanced far enough so that adequate concentrated blood derivatives were available for all patients with hemophilia in the United States.

This study was supported in part by Health Services Administration Grant MCB-360001; Health and Human Services Grant HS-30567; The Regional Comprehensive Hemophilia Diagnostic and Treatment Center; the Margie Boas Fund; the International Hemophilia Training Center of the World Federation of Hemophilia; The Polly Annenberg Levee Hematology Center, Department of Medicine, Mount Sinai School of Medicine of the City University of New York; and Grant 5MO1RR00071 for the Mount Sinai General Clinical Research Center from the National Center for Research Resources, National Institutes of Health.

From The Mary Weinfeld Professor of Clinical Research in Hemophilia, Mount Sinai School of Medicine, New York, New York.

Address reprint requests to Louis M. Aledort, M.D., Mount Sinai Medical Center, One Gustave L. Levy Place, Box 1006, New York, NY 10029.

Many insurance companies at that time, including government-sponsored ones, were beginning to recognize the benefits, both physically and fiscally, of "home care," and the comprehensive care approach to this disease. In 1975, the U.S. government passed the Hemophilia Act of 1975 which set aside funding for developing a network of hemophilia treatment centers throughout the United States. This network has grown and now cares for the majority of patients.

However, at the same time that the morbidity and mortality of this disease markedly improved, we began to recognize the problems associated with the use of human concentrates. Hepatitis B, non-A-non-B liver disease, and exposure to large amounts of foreign protein were being recognized (2). Clearly, long-term epidemiologic studies were needed and conducted.

During the late 1970s to early 1980s the life of hemophiliacs changed dramatically, with marked improvement in school attendance, reduction in time lost from work, and significant decrease in the use of the hospital, as well as the ability now to attain a normal life-span (3). These advances were altered substantially by the appearance of the HIV virus in the blood supply and the first cases of AIDS in hemophilia in 1982. This major disaster, affecting 70% of hemophiliacs in the United States, has led to major changes in patient outlook, fractionation technology, patient attitudes, and a broad information base that has changed the outlook for the current newborn who has hemophilia. The 1980s were consumed with managing HIV infection. Understanding the epidemiology of HIV, its treatment, and how blood products influence outcome has been a priority. In addition, the manner in which non-A-non-B hepatitis, now recognized as hepatitis C, affects outcome in HIV-negative and -positive patients has become a critical issue in patient care.

The current availability of both viral-inactivated and recombinant products offers the likelihood of viral-free products. Newborns may be able to look forward to a viral-free life. In addition, newer modalities of therapeutic approaches such as prophylaxis and gene therapy may allow patients with hemophilia to prevent joint disease.

Viral Transmission of Blood Products

The HIV retrovirus entered the blood supply in 1979, and by 1982 more than 50% of patients with hemophilia in this country were infected, and almost 90% by 1983 (4). However, not until the retrovirus was identified, and the epidemiologic studies by the Centers for Disease Control published in 1984, did the medical community recognize that this disease could be carried and transmitted by transfusion. By the end of 1984, investigations revealed that this virus was exquisitely heat sensitive. This led immediately to the rapid conversion of almost all available hemophilia products from nontreated to viral-inactivated products. Since 1987, no new HIV seroconversions in patients with hemophilia have been reported in this country. Techniques for viral inactivation have gone through many sophisticated laboratory changes and now include heat, vapor treatment, solvent-detergent treatment, ion-exchange chromatography, and monoclonal separation.

The other viral infection of great concern to this patient population is hepatitis C. The viral-inactivation techniques have all but eliminated the transmission of this agent as well. The agent that has caused little or no disease, but is still a concern because it may represent a marker virus for unknown agents, is Parvo B (5). Although a large percentage of healthy patients and patients with hemophilia have evidence of past exposure, only a few patients with hemophilia have had transient clinical disease related to transfusion. However, these observations have led to continued attempts to further refine inactivation techniques. The discovery of the molecular structure of factors VIII and IX has led to the ability, using current molecular biologic technology, to produce recombinant factor VIII (6), which is now licensed, and recombinant factor IX, which is currently in clinical trials. Although, theoretically, recombinant factors should be free of transmissible agents, too little time has passed to feel secure that infection will not occur either from the culture media or cells into which the DNA has been placed. In addition, all licensed recombinant VIII products are stabilized with albumin derived from human plasma, and thus from a large number of donors.

The anxiety and skepticism about the safety of the blood supply has led to continued surveillance of this system. Recent recognition that patients with Creutzfeldt-Jakob disease have donated to the plasma pool for blood derivatives has led to a recall of many products, including albumin. Although there is no evidence that this disease has ever been transmitted via blood, such materials will continue to be recalled; this adds to the concern over albumin-stabilized recombinant materials. Our government, after an Institutes of Medicine review of the evolution of HIV from blood, has implemented a new Department of Health and Human Services-driven surveillance of the blood banking and plasma derivative industry. A great deal of focus, appropriately, is on the quality of the donor as well as donor screening. New donor tests will be implemented rapidly to decrease the likelihood of transmission of disease by either eliminating or reducing the windows which exist when current tests can identify viral infection.

The Appendix has a list of all products now available for patients with hemophilia. Physicians and patients have a large armamentarium of products from which to choose, based on each patient's choice, clinical condition, and ability to pay. All these products are efficacious but, unfortunately, not all have been tested similarly for viral safety, so only some can show definitive advantages over others. A lengthy discussion of purity and safety goes beyond the scope of this paper.

Hepatitic C Virus

From the early recognition that non-A-non-B hepatitis in patients with hemophilia caused significant liver pathology (7), physicians have conducted epidemiologic studies to define its natural history. Recently it was discovered that almost all of those patients with what was described as non-A-non-B hepatitis had, in reality, hepatitis C (HCV). Using current testing methods, 100% of patients with hemophilia treated with nonvirally inactivated products were infected with that virus. The combination of HIV and HCV has been of particular concern to this patient population, as it leads to a multiplicity of clinical problems. These include difficulty in administering antiretrovirals as well as prophylaxis for opportunistic infections. Extreme liver functional abnormalities ensue, making discontinuation of medication essential. More serious, however, is the recognition that hepatitis C viral titers are significantly higher in HIV-positive versus HIV-negative individuals (8). In addition, their liver disease progresses; liver failure is now a leading cause of death in HIV-positive patients. This leads to two more coagulation defects: decrease of factors synthesized by the liver and a qualitative platelet defect. Portal hypertension with bleeding esophageal varices is a common mode of exodus. Because hepatitis C has been active in these patients for many years, many of these patients are not candidates for intervention with α-interferon. For those eligible for this therapy, no large-scale trial has been carried out to give us definitive guidelines for management, but current data do show that some patients respond with loss of abnormal enzymes. Because liver biopsy is a potential hazard for these patients, little is known regarding changes in pathology (7). In addition, some HCV titers and subtypes of hepatitis C are resistant to α-interferon (9). No systematic titers and/or typing of the virus have been carried out in patients with hemophilia so treated; thus, we cannot adequately evaluate the failures. Further studies are needed to determine the role of interferon in the management of patients who have hemophilia with hepatitis C, with or without HIV infection.

Inhibitors

Antibodies to coagulation factors, although occurring in patients without hemophilia, are most common in patients with congenital coagulation defects who have received treatment. These antibodies can be of significant titer so that the infused clotting factor is inactivated, and hemostasis may not be achieved easily. A great deal of interest has been reignited recently as to the incidence of inhibitor formation from various products, the nature of these antibodies, the predictability of which patient is likely to develop an inhibitor, and, not least important, the management of these patients with inhibitor.

A review of the literature up to the 1980s reveals that approximately 10% to 15% of patients with hemophilia who are deficient in factor VIII develop inhibitors. The majority of these occur before the age of 20 years, and most are of high-titer, anamnestic, and thus are clinically difficult to manage (10). Recent prospective studies of patients previously untreated (PUPs), and now receiving monoclonal or recombinant factor VIII, have developed inhibitors in as many as 30% of recipients of the latter products. Approximately one-half of these inhibitors are of high titer, the remainder either disappearing or remaining low titer. The surveillance of these patients far exceeded previously reported studies and, because of differences in data collection, one cannot adequately make comparisons. However, some recent publications using lower purity products indicate a lower proportion of patients develops inhibitors. Although the issue of relative incidence of inhibitors related to product purity remains unresolved, inhibitors from recombinant products occur early, after only 9 to 12 exposures, and in very young children (11).

Predisposition to inhibitor development has been studied for some time, and it is recognized that black patients have twice the likelihood of developing inhibitors as white patients. In addition, concordance for inhibitor development in families is well recognized. The search for commonality of HLA types or groups for predictability has not been successful, but the recent ability to define epitope specificity of these antibodies has led to a new avenue of investigation which may prove fruitful. Epitopes against the A_2 and C_2 portions of the factor VIII molecule have now been studied extensively in patients with hemophilia (VIII-deficient) who have developed inhibitors with low-purity products, who have been treated exclusively with recombinant products, and who do not have a congenital coagulation defect (12). Currently, although somewhat disappointingly, all groups studied show similar patterns of epitopes.

Another approach has been to begin a systematic analysis of the mutation causing the genetic defect. Deletions, missense, and nonsense mutations have all been associated with inhibitors. A new mutation, an inversion of one of the amino acid chains, has been shown to be responsible for almost 40% to 50% of the genetic defects causing hemophilia. Originally thought not to have any associated inhibitors, it is now recognized that patients with this mutation also have inhibitor induction. Prospective studies of PUPs currently underway will offer the opportunity to determine the likelihood of inhibitor development with the varied mutations. These approaches will allow us to know the statistical likelihood that an inhibitor will develop in a given patient, but not whether a given patient will develop an inhibitor. In addition, we have also recognized recently that altering fractionation processes can lead to formation of new inhibitors even in patients who have been treated in the past without evidence of inhibitors. This has occurred with a product produced in the Netherlands and one produced in Germany.

A major current thrust in treating the patient with an inhibitor is to attempt to eradicate it. This has been approached through a well-known phenomenon, immune tolerance. Using massive antigen overload coupled with activated prothrombin complex concentrates, Dr. Brachman in Bonn, Germany was able to eliminate inhibitors. These observations were not studied or replicated for many years. But as one recognized that the prevalence of inhibitor formation has not changed with higher purity, viral-inactivated, or recombinant products, various regimens have been used to reach tolerance. A spectrum of very high daily doses of factor to low thrice-weekly doses has been used. In a review of these regimens (13), it is clear that very-high-titer antibodies of longstanding are the most difficult to eradicate and frequently do not disappear. However, the rate of success is quite high for all comers. The data in the literature strongly suggest that the best results are found when intervention is early, particularly shortly after recognition, and when titers are low.

The definition of success of immune tolerance needs careful scrutiny, as its induction is lengthy, as long as 2 to 3 years, and is very costly. The optimal definition of success is that when treatment stops, and at least 3 days have passed without factor replacement, then the infusion of factor concentrate should lead to a normal retrieval and biologic half-life. In addition, this normal state should be able to be maintained without mandatory continuation of an immune tolerance regimen. No program leads to a 100% success rate. Some programs of immune tolerance succeed in changing the patient's inhibitor from high titer to low, so that factor can be used once again.

The group in Malmö, Sweden has taken a different approach which has a high success rate, shortens the time of induction, and is less costly. Using a protein A continuously eluting adsorption column, inhibitors are brought down in titer to a low level, usually less than 10 Bethesda Units. Antigen excess is used and combined with immunosuppression using cytoxan and intravenous gammaglobulin (14). This full regimen cannot be used in this country, as this protein A column is not yet available in the United States. A prospective trial of immune tolerance using antigen overload, with or without immunosuppression, is now underway.

Because the cost of immune tolerance is very high, it becomes more important to define whether or not there is a substantial difference in both lifestyle and orthopedic outcomes in patients with and without inhibitors. For at least a decade, we have recognized that treatment per se for patients with inhibitors costs more than for those without one (15). If we can demonstrate that annual costs for inhibitor patients far exceed those associated with patients without inhibitors, and their physical well-being is worse, we may be able to convince insurers that the cost of immune tolerance is worth the investment.

The management of patients with inhibitors is beyond the scope of this paper (10), but a brief commentary is warranted. Currently, only a few therapeutic agents are available as licensed products. Prothrombin complex materials can be used satisfactorily in some patients with high-titered inhibitors. The most efficacious therapy is the use of activated prothrombin complex materials (see Appendix). Although none of these agents will give one a hemostatic measurable level of factor, a large majority of bleeding episodes can be treated successfully. In some instances, factor VIII-deficient patients with high-titered human antibodies can have low cross-reactivity to porcine factor VIII, and they may get a therapeutic response to this factor (16).

Prophylaxis

Spontaneous bleeding characterizes severe hemophilia. The most common sites of bleeding are joints, with bleeds in ankles, elbows, and knees occurring with the most frequency. This, of course, is both frequent and painful, leading to substantial morbidity. Bleeding into a joint leads to damage, which progresses even without continued bleeding episodes. Thus, the ideal approach to therapy is to eliminate joint bleeding as well as to prevent the progression of disease in those joints already involved.

For many years in Sweden, Germany, and Holland, patients have been treated in a way quite different from the way hemophilia is treated in the United States. Since the advent of concentrates and home care (or self- or family-infusion), patients are treated on demand. Patients and families are instructed regarding the appropriate dose for individual joints, and they initiate treatment when bleeding occurs. Using this approach, the average patient in the United States uses about 2000 U/kg per year. Most of the patients in this country have joint disease. In the countries mentioned above, the alternative preventive approach of prophylaxis is used. Factor VIII patients receive factor three times a week and factor IX patients receive factor twice a week. Treatment is started as primary prophylaxis, usually at about 1½ years of age and before the first bleeding episode occurs. The goal of this therapeutic approach in Sweden, where most of the data have been collected, is to change the severe patient to a moderately severe patient, with a level of no less than 1.5% at all times.

This approach has now been followed for almost 25 years and has been demonstrated, in a small number of patients, to prevent any joint damage (17). In a large prospective study of orthopedic outcomes over a 5-year period, prophylaxis was superior to using factor on demand in maintaining normal joints, markedly reducing joint bleeds, and slowing the progress of joint damage in those joints already involved. This was true for both physical examination and x-ray evaluation. In addition to these findings, the study evaluated the impact of treatment regimens on patient productivity and hospitalizations. As with joint morbidity, patients on prophylaxis missed fewer days from work or school and spent fewer days in the hospital (18). Also, of note, was that patients with even rather minimal joint damage had more surgery than those with none. The more beneficial outcomes from prophylaxis cost the health care system less. On the other hand, to achieve these, the amount of factor used annually, 3000 U/kg, far exceeds what is used now on demand. The costs of factor eliminate any savings in dollars, and costs for prophylaxis still exceed those for on-demand. These costs, in our current health care environment, may make prophylaxis unfeasible (19).

Prophylaxis as a primary intervention has other issues that warrant discussion and will play a key role as to whether this treatment will become an accepted form of therapy. Most infants, by age 1½ years, do not have adequate venous access. They therefore require hospitalization to implant either a broviac or port-a-cath to gain venous access. There a substantial risk of post-implant infection which may require hospitalization, and , on occasion, removal of the device. In addition, psychologic acceptance of the three-times-a-week infusions is not always possible.

Although there is growing evidence that prophylaxis in this disease may be more beneficial than using factor on demand, no true comparative randomized prospective study has been conducted. Careful assessment of outcomes, feasibility, prevalence of venous access devices, complications of therapies, and total costs of care are essential to make appropriate clinical recommendations. Health care economists have already determined the costs of hemophilia associated with productivity as well as the costs associated with days lost from work or school, and hospitalization (5). On the other hand the financial value of psychologic benefits, such as the feeling of well-being, has not been evaluated. When these data are available, HMOs or other insurers may be sufficiently convinced to underwrite this expensive therapy.

Continuous Infusion

How a replacement regimen is decided on relies heavily on the biology of the factor. The recovery, biologic half-life, the desired factor level to be achieved, and estimated plasma volume are all key factors in determining how much factor, how often, and for how long to treat. The cost of factor replacement has always been the single most costly part of the annual management of a patient with hemophilia. Thus, conservation of factor is of great concern and interest to physicians and the health care system. For many years, when surgery has been carried out, some centers have used a technique of continuous infusion in an attempt to eliminate the peaks and valleys of factor levels resulting from repeated bolus injections. Little data existed, however, as to how much factor conservation would be achieved using this technique. This, in part, was due to the fact that for a long time, pump systems were inadequate to deliver the material at a constant volume. In addition, little or no information existed as to the retention of potency of factor over time, the avidity of factors for the surfaces to which they were exposed, the safety of maintaining the factor over many hours, i.e., bacterial contamination, and the reduction of factor requirement that could be achieved at various final constant factor levels in the patient. All these issues are receiving a great deal of attention. In addition to concerns about cost containment, ideal prophylactic programs are being sought.

This has led to studies of viability of products, both factor VIII and IX, over time at room temperature. It appears that these products are viable for considerably longer than Food and Drug Administration licensure allows. New highly concentrated products also allow for small volumes to be used to deliver factor over long periods of time. Many surfaces are avid for factor VIII, and the addition of albumin may be necessary to prevent substantial losses. New devices are being evaluated to determine an ideal one to deliver a constant volume over long periods. These still require long-term venous access, and all of the issues stated above still pertain. However, at present, no single licensed product has been accepted by the Food and Drug Administration as a safe and effective therapy for longer than currently accepted length of time at room temperature as stated in the package insert. Fractionation companies are now studying these issues in an attempt to get a change in indication. In addition, companies are investigating techniques to deliver small-volume, high-concentrate materials exposed to surfaces that do not adsorb a significant amount of factor.

Gene Therapy

The advent of the isolation and structure of the DNA molecule for factors VIII and IX has made possible the insertion of these genes, which produce factors VIII and IX, into cells. This technique has led to the successful production of recombinant factors VIII and IX. Another result has been the ability to experiment with these for gene therapy for these two hemophilias. The work on gene therapy for a variety of diseases was initiated in 1989, and currently more than 100 clinical trials are being undertaken. The bulk of the diseases being attacked are cancer and AIDS, with a small but substantial amount of effort directed toward hemophilia (20).

The early work used retroviruses, adenoviruses, and adeno-associated (parvo) viruses as vectors for therapeutic gene delivery. Some nonviral delivery systems have included plasmas and liposome-DNA complexes. The efforts for factor VIII and IX gene therapy have been associated with either difficulty in achieving a significant level of factor or the inability to sustain a level. An ideal vector would be one capable of achieving a level high enough to prevent spontaneous bleeding, would not be needed to be administered too often, and would not be infectious.

Little work has been carried out with factor VIII because of the size of the molecule. Most effort has focused on the factor IX molecule, using a dog hemophilia B model. Adeno- and retroviral vectors have been used. The retroviral vectors produce longer-term factor levels but with low expression. The adenoviral vectors can achieve high levels but are short lived. Liver cells, cardiac, and skeletal muscle cells have all been used. Inhibitor induction to both vector and final product also plays a role in limiting the efficacy of gene therapy. Although patients and physicians looked to gene therapy as a "cure" for hemophilia, the experimental results are disappointing. However, new techniques of investigation, advances in molecular biology, and innovative vectors and target cells may substantially alter this field. Immediate answers are less likely to be realized, but a longer-term perspective should be optimistic (21).

Financing Care

Throughout this paper, I refer to the cost of care and the challenges that newer modalities of treatment may make on the health care system. In most of the Western world, health care is a covered entity for most citizens, but it is not in the United States. We live with a pluralistic health care system with almost 40 million uninsured persons. For those who are insured, few with a chronic disease have full coverage for their medical, surgical, as well as treatment costs. For hemophilia, the majority of costs are those for the factor concentrate. At best, patients have insurance that covers about 80% of these costs with a lifetime cap of, usually, one million dollars.

Medicare will cover 80% of those costs that they consider fair and published in the Federal Register. Medicaid varies from state to state and city to city. For New York City, we have been able to convince Medicaid to cover 100% of the blood products used at home, and the coverage is annually renewable. This is also true for some subscribers with Blue Cross and Blue Shield.

The issue facing many hospitals, such as Mount Sinai, where there has been a longstanding comprehensive hemophilia clinic, is that many patients in the region use the hospital, not only for their out-patient clinic care, but for emergency room and in-hospital care. With diagnostic-related groups, a flat rate is negotiated for a hospital stay and for emergency room visits. These rates do not adequately cover the cost of blood product, which is almost always required for these encounters. Medicare, for the past few years, has had a codicil and paid diagnostic-related group plus product. With the current stalemate in Washington regarding the budget, this entitlement no longer exists.

The greatest threat to hemophilia care today is the proliferation of managed care. Even with its early penetration in the New York area, we have already seen the problems facing the patients with this chronic disease. Access to the annual comprehensive evaluation is denied, as the insurer does not perceive its need. Special laboratory tests such as factor assays and inhibitor titers are refused, as they cost more than with laboratories with which the insurers have contracts. It is an uphill fight to convince them of the lack of quality control and accuracy of the assay at many of these laboratories. Far more serious is the conviction by some HMOs that their primary care physicians are as competent in managing hemophilia bleeding as those experts who have devoted themselves to this disease. The questioning of the quality of care provided by HMOs is just beginning to occur. The federal government has provided substantial resources to develop centers of excellence for low-density, life-long genetic diseases. These specialized centers of excellence should not be dismantled, nor should these patients lose these services.

The Mount Sinai hemophilia program is one of the leading health care systems for these patients in the United States and has been designated an International Hemophilia Training Center by the World Federation of Hemophilia. It has provided guidelines which define comprehensive care and the issues involved in its financing.

Conclusion

Hemophilia is an old disease. The 1940s brought the science of coagulation together with modern blood-banking technology. This led to marked improvement in therapy, novel delivery of health care, and a major change in outlook for these patients. The model of their care has served as a template for other chronic diseases. The advent of HIV and HCV caused a dramatic upheaval. However, this tragedy has led to innovations in fractionation, and advances in molecular biology have made recombinant therapy a reality and gene therapy a future possibility. Novel health care delivery has established prophylaxis and continuous infusion as future therapies. The continued presence of inhibitors in treated patients, and their eradication, offers a substantial challenge for therapy. All these advances, from safer products to newer interventions, are more costly, but offer our patients, especially newborns, a bright future. Whether our nation's commitment to health care can provide the resources for their future remains to be seen.

APPENDIX

HHS Region II Hemophilia Education Task Force
Hemophilia Treatment Products Reference Guide

This is a practical guide for educational purposes and does not reflect all indications for use. The information contained in this guide was compiled from product package inserts. It should be noted that various manufacturing methods are utilized in the production of clotting factor concentrates which result in varying relative product purity and viral safety. For further information, consult prescribing physician, product package insert or Hemophilia Treatment Center.

Factor VIII Products (Plasma-Derived)

Product name Manufacturer	Description	Rate of infusion	Half-life	Action	Reconstituted use within	Undiluted storage
AHF-M METHOD M Baxter/Hyland for Am. Red Cross	monoclonal; solvent detergent	10 ml/min	14.8 ± 3 hr	1 unit of factor ↑ plasma level 2%	1 hr	refrig 2–8°C
ALPHANATE Alpha	solvent detergent	not to exceed 10 ml/min	17.9 ± 9.6 hr	1 unit of factor ↑ plasma level 2.4 ± .4%	3 hr	refrig 2–8°C or room temp × 2 months
FVIII SD NYBC	solvent detergent	not stated	no information given	1 unit of factor ↑ plasma level 2–2.5%	3 hr	refrig 2–8°C
HEMOFIL-M Baxter/Hyland	monoclonal; solvent detebgent	10 ml/min	14.8 ± 3 hr	1 unit of factor ↑ plasma level 2%	1 hr	refrig 2–8°C
HUMATE-P Armour	pasteurized	4 ml/min	12 hr	1 unit of factor ↑ plasma level 2–2.5%	3 hr	refrig 2–8°C or room temp × 6 months
KOATE-HP Cutter/Miles	solvent detergent	5–10 min	9–14 hr	1 unit of factor ↑ plasma level 2–2.5%	3 hr	refrig 2–8°C or room temp × 6 months
MELATE NYBC	solvent detergent	10 ml/min	10 hr	1 unit of factor ↑ plasma level 2–2.5%	3 hr	refrig 2–8°C
MONOCLATE-P Armour	monoclonal; pasteurized	2 ml/min	17.5 hr	1 unit of factor ↑ plasma level 2%	3 hr	refrig 2–8°C or room temp × 6 months
PROFILATE-OSD Alpha	solvent detergent	10 ml/min	12 hr	1 unit of factor ↑ plasma level 2–2.5%	3 hr	refrig 2–8°C or room temp × 3 months

Factor VIII Products (Recombinant)

Product name Manufacturer	Description	Rate of infusion	Half-life	Action	Reconstituted use within	Undiluted storage
BIOCLATE distrib. by Armour manufactured by Baxter	recombinant; glycoprotein synthesized by genetically engineered chinese hamster ovary cell line; purified using a series of chromatography columns	10 ml/min	14.4 ± 4.9 hr	1 unit of factor ↑ plasma level 2%	3 hr	refrig. 2–8°C or room temp (not over 86°F)
HELIXATE distrib. by Armour manufactured by Miles	recombinant; monoclonal antibody immunoaffinity chromatography & lyophilization	over 5 min	15 hr	1 unit of factor ↑ plasma level 2%	3 hr	refrig. 2–8°C or room temp × 3 months
KOGENATE Miles	recombinant; monoclonal antibody immunoaffinity chromatography & lyophilization	over 5 min	15 hr	1 unit of factor ↑ plasma level 2%	3 hr	refrig. 2–8°C or room temp × 3 months
RECOMBINATE Baxter/Hyland	recombinant; glycoprotein synthesized by genetically engineered chinese hamster ovary cell line; purified using a series of chromatography columns	10 ml/min	14.4 ± 4.9 hr	1 unit of factor ↑ plasma level 2%	3 hr	refrig. 2–8°C or room temp (not over 86°F)

Factor IX Products

Product name Manufacturer	Description	Rate of infusion	Half-life	Action	Reconstituted use within	Undiluted storage
ALPHANINE** Alpha	heated in *n*-heptane solution	10 ml/min	21 hr	1 unit of factor ↑ plasma level 1%	3 hr	refrig 2–8°C
ALPHANINE SD** Alpha	solvent detergent	10 ml/min	21 hr	1 unit of factor ↑ plasma level 1%	3 hr	refrig 2–8°C
BEBULIN* Immuno	vapor heating	2 ml/min	20–24 hr	1 unit of factor ↑ plasma level 1–2%	3 hr	refrig 2–8°C
KONYNE-80* Cutter/Miles	dry heat	100 U/min	24 hr	1 unit of factor ↑ plasma level 1%	3 hr	refrig 2–8°C or room temp × 1 month

Factor IX Products

Product name Manufacturer	Description	Rate of infusion	Half-life	Action	Reconstituted use within	Undiluted storage
Mononine** Armour	monoclonal antibody immunoaffinity chromatography; chemical treatment	2 ml/min	24–30 hr	1 unit of factor ↑ plasma level 1%	3 hr	refrig 2–8°C or room temp × 6 months
PROFILNINE HT* Alpha	heated in *n*-heptane solution	10 ml/min	17–32 hr	1 unit of factor ↑ plasma level 1%	3 hr	refrig 2–8°C or room temp × 1 month
PROPLEX-T* Baxter-Hyland	dry heat	3 ml/min	24–32 hr	1 unit of factor ↑ plasma level 1%	3 hr	refrig 2–8°C

* Prothrombin complex concentrate (contains varying amounts of factors II, VII, IX, and X).
** Coagulation Factor IX concentrate.

Factor VIII Inhibitor Products

Product name Manufacturer	Description	Rate of infusion	Half-life	Dosage	Reconstituted use within	Undiluted storage
AUTOPLEX-T* Baxter/Hyland	heat treated	2 ml/min	not stated	25–100 U/kg body weight	3 hr	refrig 2–8°C
FEIBA-VH* Immuno	vapor treated	2 U/kg/min	not stated	25–100 U/kg body weight; not to exceed 200 U/kg/ 24 hr	3 hr	refrig 2–8°C
HYATE-C Porton (porcine AHF)	polyelectrolyte chromatography purification	2–5 ml/min	no half life given	dose dependent upon FVIII inhibitor level	3 hr	frozen −15° to −20°C

NOTE: Use these products only in patient with inhibitors to Factor VIII.
* Complex of Vitamin K-dependent factors.

Factor VIII and von Willebrand's Disease

Product name Manufacturer	Product information	Rate of infusion	Dosage/ action	Reconstituted use within	Undiluted storage
CRYOPRECIPITATE	1. 1 bag contains approx. 100 units of factor VIII 2. Used for persons with Factor VIII deficiency and VonWillebrand's disease 3. This product is not viral-inactivated	Infuse within 1 hr	1 cc ↑ plasma level 2%	thawed: 4 hr if pooled, 6 hr if not pooled	frozen
DDAVP (Desmopressin Acetate) Armour	1. Increases factor VIII activity and VWF levels for patients with mild to moderate disease. 2. Dose is diluted in 50 ml of normal saline for adults and children weighing more than 10 kg; in children weighing 10 kg or less, 10 ml of diluent is used. 3. Facial flushing is a usual side effect. 4. Possible side effects—transient headache (often rate-related), slight rise in blood pressure, nausea. 5. Response decreases if doses are repeated in less than 48 hours. 6. Should not be used in VWD type IIB. 7. This is a medication—NOT a blood product	Infuse slowly over 15–30 min	0.3 µg/kg body weight	24 hr	refrig
STIMATE NASAL SPRAY (Desmopressin Acetate) Armour	1. Same action and indication as DDAVP (intravenous) 2. Intranasal Stimate provides less factor rise than intravenous DDAVP. 3. Lab tests indicated for trial dose to determine if factor levels are clinically adequate to provide hemostasis. 4. Peak plasma concentration 40–50 min after dosing; half-life is 3–3½ hr 5. Tachyphylaxis can occur if dosing is repeated within 48 hr 6. Same side effects as DDAVP. 7. Packaged as a 5 ml intranasal spray pump bottle—2.5 ml of medication in each bottle. 8. Each bottle contains 25 doses of desmopressin—150 µg (1.5 mg) of each dose.	N/A	Dose by body weight: <50 kg—give 1 spray into nostril (150 µg) >50 kg— give 2 sprays, one into each nostril (total 300 µg)	N/A	Store under refrigeration

Factor VIII and von Willebrand's Disease

Product name Manufacturer	Product information	Rate of infusion	Dosage/action	Reconstituted use within	Undiluted storage
HUMATE-P* Armour	pasteurized factor VIII concentrate contains VW protein half life—12 hours	4 ml/min	no information given	3 hr	refrig 2–8°C or room temp × 6 months

* Though not currently licensed for use in VonWillebrand's disease (VWD), Humate-P has been used in management of bleeding episodes associated with VWD. Contact Hemophilia Treatment Center or prescribing physician for more information.

Any Factor Deficiency

Product name Manufacturer	Product information	Rate of infusion	Dosage	Reconstituted use within	Undiluted storage
FRESH FROZEN Plasma (FFP)	1. 1 cc contains 1 unit of all clotting factors. 2. Can be used for Factors VIII or Factor IX deficient patients and any other single factor deficiency. 3. This product is not viral-inactivated.	Administer unit within 1 hr	Children 10 ml/kg body weight; Adult dose—limited by volume, may cause fluid overload	refrig good for 6 hr after thawing	frozen

Antifibrinolytic Agents

Product name Manufacturer	Product information	Rate of infusion	Dosage	Reconstituted use within	Undiluted storage
AMICAR (Aminocaproic acid) Lederle	1. Systemic antifibrinolytic 2. Useful in helping to control mucous membrane bleeding 3. Available in tablets, syrup and intravenous formula	Variable, refer to product information	50–100 mg/kg body weight every 6 hr	24 hr	room temp
CYKLOKAPRON (tranexamic acid) Kabi Vitrum	1. Systemic antifibrinolytic 2. Available in tablets and intravenous formula 3. Useful in controlling mucous membrane bleeding 4. In patients with renal insufficiency, see package insert for decreased dosage	Refer to product information	IV 10 mg/kg body weight every 8 hr; PO 25 mg/kg body weight 3 to 4 times a day	24 hr	room temp

Topical Treatments

Product name Manufacturer	Product information	Form	Dosage	Reconstituted use within	Undiluted storage
SURGICEL Johnson & Johnson	1. Supplied in form of sterile knitted cellulose fabric as an absorbable hemostat 2. Used adjunctively for control of capillary, venous and small arterial hemorrhage	Topical	n/a	n/a	room temp
TOPICAL THROMBIN THROMBOGEN Johnson & Johnson THROMBOSTAT Parke Davis	1. Clots fibrinogen directly 2. Topical use only for accessible capillaries and small venules 3. Supplied in form of powder, solution or spray	Topical	Available doses—1000, 5000, 10,000, and 20,000 unit vials	3 hr refrigerated	2–8°C

Factor concentrate is packaged with the bottle of lyophilized factor, a bottle of diluent, one double-ended needle, and one filter needle.

1. Place diluent bottle in warm water, unless it is stored at room temperature.
2. Put on gloves.
3. Remove caps from diluent and factor vials and wipe stoppers with alcohol swab.
4. Add diluent to factor using double-ended needle:
   A. Remove needle cover from short end and insert into diluent vial.
   B. Remove cover from long end of needle; invert diluent vial and insert exposed needle into factor vial, allowing vacuum to draw diluent into factor.
   C. Remove diluent vial, leaving needle in the factor vial to eliminate any remaining vacuum.
   D. Carefully remove the needle from the factor vial and discard.
5. *Gently* rotate the factor vials until factor is completely dissolved. Do not shake vigorously.
6. Attach filter needle to syringe.
7. Unless filter needle provided has air side port, inject air into factor vial equivalent to the amount of diluent and withdraw dissolved factor.
8. Remove filter needle from syringe; do not inject air bubbles from syringe into the air.
9. Attach syringe to butterfly needle, infusing Factor at rate recommended by manufacturer.

Additional Information

- After reconstitution, keep factor concentrate at room temperature—*DO NOT* refrigerate.
- Do not waste factor. Use the whole vial. Treating with 50–150 U higher or lower than the calculated dosage is permissible.
- For more information, refer to the package insert.
- Reconstitution and usage time refers to sterility when mixed under conventional methods. For duration of efficacy, refer to individual package insert. Product stability may be increased if reconstituted under sterile conditions, thus allowing for continuous drip infusion. Consult product information.

References

1. Gilbert M, Aledort LM eds: Comprehensive care in hemophilia: A team approach. Mt Sinai J Med 1977; 44:313–479.
2. Fratantoni JC, Aronson DL, eds: Unsolved therapeutic problems in hemophilia. Bethesda, MD: US Dept of Health, Education and Welfare publication No. (NIH)77-1089:1-132, 1976.
3. Aledort LM. Lessons from hemophilia [editorial]. N Engl J Med 1982; 306:607–608.
4. Kroner B, Rosenberg PS, Aledort LM, Alvord WG, Goedert JJ for the Multicenter Hemophilia Cohort Study. HIV-1 infection incidence among persons with hemophilia in the United States and Western Europe, 1978–1990. J Acquir Immune Defic Syndr 1994; 7(3):279–285.
5. Ross-Degnan D, Soumerai SB, Avorn J, Bohn RL, Bright R, Aledort LM. Hemophilia home treatment: Economic analysis and implications for health policy. Int J Technol Assess Health Care 1995; 11(2):327–344.
6. Schwartz RS, Abildgaard CF, Aledort LM, et al., Recombinant Factor VIII Study Group. Human recombinant DNA-derived antihemophilic factor (factor VIII) in the treatment of hemophilia A. N Engl J Med 1990; 323:1800–1805.
7. Aledort LM, Levine PH, Hilgartner M, et al. A study of liver biopsies and liver disease among hemophiliacs. Blood 1985; 66:367–372.
8. Eyster ME, Fried MW, Di Bisceglie AM, Goedert JJ for the Multicenter Hemophilia Cohort Study. Increasing hepatitis C virus RNA levels in hemophiliacs: Relationship to human immunodeficiency virus infection and liver disease. Blood 1994; 84(4):1020–1023.
9. Alter HJ. To C or not to C: These are the questions. Blood 1995; 85:1681–1695.
10. Aledort LM. Inhibitors in hemophilia patients: current status and management—study in the development of inhibitor antibodies. Am J Hematol 1994; 47:3:208–217.
11. Bray GL, Gomperts ED, Courter S, Gruppo R, et al. and the Recombinate Study Group. A multi-center study of recombinant factor VIII (recombinate): Safety, efficacy and inhibitor risk in previously untreated patients with hemophilia A. Blood 1994; 83(9):2428–2435.
12. Hoyer LW. Characterization of factor VIII inhibitors. In: Recent developments in the study of inhibitors to factor VIII therapy. American Red Cross 1994:5–8.
13. Mariani G, Ghirardini A, Bellocco R. Immune tolerance in hemophilia: Principal results from the international registry: Report of the factor VIII and IX subcommittee. Thromb Hemostasis 1994; 72(1):155–158.
14. Nilsson IM, Berntorp E, Zettervall O. Induction of immune tolerance in patients with hemophilia and antibodies to factor VIII by combined treatment with intravenous IgG, cyclophosphamide, and factor VIII. N Engl J Med 1988; 318:947–950.
15. Aledort LM, Cohen M, Hilgartner MW, Lipton R. Treatment of hemophiliacs with inhibitors: Cost and effect on blood resources. In: Hoyer L, ed. Factor VIII inhibitors. New York: Alan R. Liss, Inc., 1984:353–365.
16. Aledort LM, ed. Factor VIII inhibitor treatment: Strategies for the continuing challenge. Semin Hematol 1994; 31(2) (suppl 4).
17. Nilsson IM, Berntorp E, Lofquist T, Patterson H. Twenty-five years' experience of prophylactic treatment in severe haemophilia A and B. J Intern Med 1992; 232:25–32.
18. Aledort LM, Haschemeyer RH, Pettersson H. A longitudinal study of orthopedic outcomes for severe factor VIII deficient hemophiliacs. J Intern Med 1994; 236:391–399.
19. Bohn RL, Avorn J, Glynn RJ, Choodnovskiy I, Haschemeyer R, Aledort LM. Prophylactic use of factor VIII: Are the clinical benefits worth the additional cost? Blood Abstr 1995; 286(10):611a.
20. Blaese RM. Steps toward gene therapy. 1: The initial trials. Hosp Prat 1995; 30(11):33–40.
21. Hoots WK. Future aspects of hemophilia treatment. Biomed Progr 1995; 8:37–41.

Prevention into the 21st Century

ARTHUR L. FRANK, M.D., PH.D.

Abstract

Prevention in the United States has suffered because of the use of a traditional "medical" model rather than a "health" model. Financing, medical education, and research support have favored a focus on disease rather than prevention, although the greatest gains in man's physical well-being have come from preventive rather than curative activities. Although prevention has not been a major focus within medicine to date, despite its significant contributions, it appears that major changes will take place in the twenty-first century that will markedly change this situation. Mount Sinai, because of its meaningful role in studying the health of groups and in health service delivery, and its founding as a medical school with a focus on populations, has made, and will make, contributions to preventive activities in the future.

BECAUSE THIS IS WRITTEN IN HONOR of the centenary of the Alumni Association of Mount Sinai, an institution with which the author has been closely and proudly associated, the solidity of such an organization and institution stands in sharp contrast to the "shifting sands" of medicine at the close of the twentieth century. The field of Preventive Medicine has a long and, in many ways, distinguished history. Although currently suffering in ways not seen before, its prospects are bright in contrast to the curative specialities in medicine. The twenty-first century surely will be better for prevention, but not without much pain and suffering for both patients and medical professionals.

Since the beginning of time, one of the biggest risks to most humans was dying early in life from an infectious agent. Later life often carried similar perils, but the greatest contributions to the extension of life in man, most of it coming in the twentieth century, has been the control of infectious agents. In the space of less than 100 years the focus in advanced societies has shifted dramatically from major concerns about infectious agents, through concerns about chronic disease, and to the current view of trying to control what might be broadly characterized as environmental diseases. Unfortunately, for most of the peoples of the earth, they are still mired in the era of infectious hazards, and those in the "advanced" world may soon be rejoining them (1).

The reason for our current dilemma is founded in two fundamental problems, the second stemming directly from the first. Rather than adopting a world view of the ancient Chinese who paid their physicians for each day that they were well, the "health model," our society embraces instead the "medical model," which treats after the development of illness. This skewed view of the human condition has lead to the second, very real dilemma: the lack of true appreciation of the health needs of people and the minimal amount of support ordinarily provided. Although it is disquieting that some 40 million Americans presently have no health insurance, and millions more are underinsured, this still addresses only the medical model of caring for individuals when they become sick.

Equally if not more devastating than the lack of insurance for many is the trend of diminishing support for public health—little wonder that the Institute of Medicine has recently documented the disarray of the public health infrastructure in

---

From the Office of Medical Education, The University of Texas Health Center at Tyler. Address reprint requests to Dr. Arthur L. Frank, Topperman Professor of Medical Education, University of Texas Health Center at Tyler, P.O. Box 2003, Tyler, TX 75710.

the United States. The current restructuring of medical care in the United States is doing little to reverse this situation and, in many ways, is aggravating the problem as profitability of health care becomes of paramount importance.

Americans have never been known to favor the "collective good" view of life; "rugged individualism" has generally been more valued, with pride for those who succeed and rise above the pack. To a large extent, we have even lost interest in how our differences give us strength, and great value is often placed on behavior and attitudes formerly unwelcome in our society. It might be helpful to examine briefly how public health has evolved in recent years, assess its contributions and failures, and look to the future.

Evolution of Public Health and Prevention

As noted above, prevention activities focused primarily on dealing with infectious agents. The provision of clean water and food was of high concern, followed by collection of waste. At the time The Mount Sinai Hospital was founded in the middle of the 19th century, much of New York still was a true cesspool of human and animal wastes, dead horse carcasses, and overcrowded living conditions, with all that this implied. The twentieth century still saw thousands of cases of diet-related diseases, tuberculosis, and rampant malaria. Hospitals were generally thought of as places where people went to die, an image that has changed only relevantly recently. Hospitals have also evolved into the setting for the central focus of care for many situations previously dealt with at home, such as childbirth and long-term recovery from strokes and other such problems. Ironically now, due to the financial pressure in the health care arena, we are "forcing" women to return home soon after normal delivery, rather than staying in hospitals. Professional societies are now challenging the medical payors by stating that a 2-day stay in the hospital is not inappropriate. This is not the forum to decide where, in fact, some types of health care would be optimally delivered, but one can observe that current debates have little to do with the broader view of what ultimately is best for the public or patient, rather than the pocketbooks of investors in health care businesses, or the personal well-being of physicians.

With notable success in dealing with many infectious problems, especially in economically developed countries, the problems of societal concern turned from early death from infectious causes to the problems of chronic illnesses such as heart disease, the consequences of hypertension, diabetes, and similar issues. That overall concern was relatively short lived, perhaps only a generation or two, because rapid progress was made in caring for those with such afflictions. One could now, for example, easily, if not expensively, clean clogged arteries around the heart or in the neck. Prevention of the underlying problem was not yet in vogue, but more of that later.

There are clearly a number of reasons with strong historical roots why prevention is not yet in great favor. Much of the difficulty comes from traditional organized medicine, but medicine itself has created much of its own dilemma. Notwithstanding the work of Pasteur, Semmelweiss, and Salk, few great advances in medicine are commonly identified as coming from preventive activities. Advances in the area of prevention are often made by groups, rather than individuals, making instant recognition more difficult.

More than anyone else, Abraham Flexner influenced for decades the model of medical education in America (2). Central to his work was the creation of organized, well-structured, high-quality educational experiences for students of medicine. In addition, this medical model had to provide comfort to those already stricken. His writings do not address the inclusion of preventive activities with medicine, or anything about keeping people well. Over time this has lead to a lack of exposure to preventive medicine, as a field within medicine, at the level of medical education where the influence on the future physicians might be greatest. At the same time, much of the knowledge and training base was not being taught in medical schools. Consequently, it came to pass that a competing educational environment was set up that was not primarily designed for physicians. In fact it was often difficult for physicians to obtain this information because Schools of Public Health, as they exist in the United States, are not primarily designed to educate physicians. Therefore, there is no strong advocate for the message of the health model.

Coupled with this lack of educational input, the other valued arena in which physicians often operate, research, has been poorly funded with regard to prevention and rarely can one make their mark here. The total eradication of a major disease, as occurred on a global scale with smallpox, is not celebrated as a triumph of medicine, but is a footnote in history with presumed little practical significance. The current resurgence of tuberculosis should remind us, however, of the impermanence to solutions of difficult problems.

When the vaccine for AIDS prevention finally comes, as it will, it will be hailed as a triumph of "medicine," not of "prevention." This is all notwithstanding the underappreciated role of organizations, such as the Centers for Disease Control and Prevention, that have done much to eradicate disease and continues to contribute to our understanding of newer diseases such as HIV and Ebola.

The lack of educational awareness and lack of a strong, identifiable research base easily translates into a lack of recognition for those who work in the field. This undervaluing of prevention can be seen in the poor individual support of its practitioners, who have the lowest salary levels of any boarded medical specialty, and in the traditionally poor support by many medical schools. For example, it is inconceivable for a medical school to be accredited without a department of surgery, pediatrics, etc., but many accredited schools have no programs in preventive medicine. In some cases such programs exist only as poorly funded divisions in larger, medically oriented departments.

Preventive medicine, as a discipline, is at fault as well for some of its perceived shortcomings. Besides the duality of medical education and public health education, there is no one easily recognized term which delineates the field. There are clearly aspects of the field that set it apart from every other specialty in medicine: It is the only field focusing on populations, rather than individuals; it's tools are different, the computer rather than the stethoscope; and the knowledge base (epidemiology, health care organization, biostatistics, etc.) is unfamiliar to those accustomed to the medical model. (This may be changing, at least with regard to the organization and financing of medical care.) Compounding this unfamiliarity is the fact that no one term has been generally agreed upon to describe the field. Although perhaps having some difference in emphasis, preventive medicine, social medicine, environmental medicine, community medicine, and the various name combinations all focus on the study of population groups.

The voice of this specialty rarely rises above a whisper. Many medical students and physicians do not even know that among the thirty or so recognized medical specialities there even exists the boarded specialty of preventive medicine. Complicating matters further is the fact that while other fields have the parent specialty and subspecialties, preventive medicine is one specialty in which certification is granted in one of several areas. Over the years general preventive medicine and public health have fused into one, rather than two areas. Joining them are aerospace medicine and occupational medicine. Although each of these areas shares common core elements, they all have significant distinctions leading to certification only in one part of the field. Interestingly, the specialty of preventive medicine is the only specialty to require significant additional academic training beyond medical school (usually culminating in an advanced degree such as M.P.H., M.S.P.H., etc.), rather than being only a carry-over from apprenticeship-like training characteristic of every other specialty.

The Economics of Medical Care

This is not the setting in which to expound either on the financial issues that are now driving medicine or on the nature of medical practice. The issue of economics is raised only to highlight, in yet another way, how differently the cost of prevention has been perceived traditionally as compared with the cost of "curative" medicine. Medicare required no scientific study of the efficacy of cardiac by-pass surgery before it began routinely paying out large sums of money for such services. Nevertheless, it took significant research and, literally, an Act of Congress, to allow medical providers to give senior citizens, under Medicare, a flu vaccination at six dollars a shot. The cost savings in terms of morbidity or cost of years saved from cardiac surgery is still unknown; the tens of thousands of dollars in morbidity savings from flu shots is well documented. Unfortunately, there are far too many examples of the shortsightedness of our society in terms of the cost savings and advancement of the overall social good (routine childhood immunization, fluoride treatments, head start programs, etc.).

The point made earlier about environmental factors being of paramount importance today is dramatically borne out by looking at the causes of years of life lost. The greatest contributors to this negative number come from dietary patterns, cigarette smoking, injury-related deaths (highway, workplace, etc.), homicide, and suicide. Complicating matters even further is the globalization of our planet. Diseases usually confined to one geographic area easily circle the globe, both infectious and lifestyle related. Poor life-style factors (improper diets, smoking) are being inflicted on previously unexposed populations. In the name of increasing profit, hazardous working conditions are being exported from developed to developing countries (3, 4). The eroding of the environment, the destruction of forests, and the reduced quality of air and water resources do not stop at geographic or political boundaries. We are not always sharing the best of what America has to offer, and our own ethnocentric views inhibit us from adopting what may be good about health care or environmental controls in other countries.

The American Prevention Scorecard

How have we fared to date in America in meeting the challenges of keeping our citizens well? Overall, we have fared quite badly. We are not the healthiest people in the world, although we spend the most on medical care. We do not lead in any of the significant indicators of the health of the public such as longevity, infant mortality, or immunization rates. Many third-world countries surpass our rates for some of these indicators.

As the last decade of the twentieth century opened there was published, after years of discussion and review, a government document which was a blueprint for improving the health of Americans. An update of Healthy People 2000 (5) outlined not generic issues, but a whole series of specific, item-by-item areas with current rates of problems and goals for the end of the century. Every one of the dozens of points had a goal for change for the better.

The current middecade report card reveals that our performance has been dreadful. Some of the problems actually have worsened. Few, if any, of the goals are expected to be achieved fully by the end of the century.

We appear, as a country, to have no will, interest, or desire to significantly improve our lot, at least not with regard to our health. We continue to eat too much of the wrong foods, with some 60% of the population overweight. We smoke too much, drink too much, use too many illicit drugs, and drive too carelessly. We raise speed limits and repeal helmet laws, and in the face of an epidemic of violent death blithely go about legalizing the use of concealed firearms. Workplace injuries and deaths continue to mount. As a nation, we do not seem to value our health, but want to continue to pay, or have paid for us, for the care needed, or desired, when we get sick.

The Prospects for the 21st Century

So far, this review has had a depressing tone. But it must be recognized that this view cannot prevail. Preventive medicine physicians are, at their core, an optimistic lot secure in their understanding of what they do, and perhaps even more comfortable with their roles than most other physicians. They as a group remain true to the ideals that traditionally have brought many people into medicine in the first place; they can say that their reward comes from helping others. This aggregate help goes far beyond the often elusive satisfaction of caring for individual patients with a possibly preventable chronic disease who will never recover from their problem.

The world of medicine will still face years of disquietude and disruption with all the push and pull of the "marketplace." Profits will be made and lost, corporate CEOs will prosper and physicians' salaries will continue to decline, having done so in the aggregate for the first time in recent memory. It will still require major changes to fully insure, at even some minimal level, all Americans. Rationing and unnecessary suffering will continue. All this will be distressing to the preventive medicine community but, in truth, they are too powerless to change things significantly.

After all the pushing and pulling and reorganizing and profit-taking within the traditional medical model, however, it will *finally* become clear that, quite simply put, prevention is the way to go. The first significant stirring of this, now to use a current buzzword, paradigm shift, is beginning to take place. The "Blues" are putting up billboards extolling the virtues of exercise and eating healthfully, insurance premiums are reflecting the costs of smoking, major efforts at improving vaccination rates are being seen, and new education models are being tried.

One such new model was identified decades ago when it was recognized that broadly based community level education in the area of health was achievable, paralleling what had been done for many years in agriculture. All of the United States shares a program of agricultural extension that covers every county in America and works at a local level, meeting the needs of everyone who has questions in relation to things agricultural. This system covers not only rural areas, but cities and inner-city areas, as well. A relative newcomer to the world of medical education, but an institution long steeped in this agricultural extension model, Texas A & M in College Station in rural Texas, has recently created what is envisioned as a rural public health program to disseminate health information in a manner that parallels their longstanding sharing of farming and ranching information through agricultural extension. It is clearly a model well worth trying.

Given the ingenuity of the American mind, and the ultimate desire to reduce the cost of medical care in this country, primarily for financial rather than humanitarian reasons, the health model will come into being as it has never existed before. It will not come easily or quickly, but with certainty, it is coming. Although one might wish for it to occur for the right reasons, those in the preventive medicine community will ultimately celebrate the achievement, even if they are unhappy about the path that will have brought us there. This is the promise awaiting medicine in the twenty-first century.

Mount Sinai, as an institution, has not been unwavering in supporting this goal. But, as an institution, it can take more pride than most other great institutions when this change finally occurs. As a hospital it was started to meet the needs of one special, underserved group in its community, and quickly expanded its role to take care of all. As a medical school it was founded on the tripod concept of basic sciences, clinical medicine, *and* community medicine. For this it can take justifiable pride. As the second century of alumni activities begins, we *will* see a stronger and better appreciated role for prevention. Undoubtedly the contributions that Mount Sinai makes toward that goal will hasten that change without loss of either its long-time luster as a preeminent clinical institution or its more recent reputation in basic biomedical research. Those who can be counted among its alumni and have been trained at Mount Sinai in the third leg of the tripod of medical education will make and leave their mark in the 21st century.

Acknowledgment

Although the thoughts expressed here have been those of the author alone, recognition of the great influence of Dr. Kurt W. Deuschle, first chair of the Department of Community Medicine at Mount Sinai, on the author and many in the field should be acknowledged.

References

1. LeCue JW. World Health Organization strategy for emerging infectious diseases. JAMA 1996; 275:318–320.
2. Flexner A. Medical education in the United States and Canada: A report to the Carnegie Foundation for the advancement of teaching, Bulletin No. 4, 1910.
3. Frank AL. Global problems from exposure to asbestos. Environ Health Perspect 1993; 101(Suppl 3):165–167.
4. Frank AL. Failure of prevention in the workplace and the added outrage of "revisionist" science. In: Mehlman MA, Upton A, eds. The identification and control of environmental and occupational diseases, Vol. XXIII, Advances in Modern Environmental Toxicology. Princeton: Princeton Scientific Publishing Co., Inc., 1994; 549–557.
5. Healthy people 2000 midcourse review. Washington, DC: U.S. Department of Health and Human Services, 1995.

Using Standardized Patients for Assessing Clinical Performance:

An Overview

Mark H. Swartz, M.D. and Jerry A. Colliver, Ph.D.

Abstract

Standardized patients are being used increasingly worldwide to assess the clinical skills and competencies of medical students, residents, and physicians; plans are now underway for the use of standardized patients in licensure and certification examinations. The Morchand Center for Clinical Competence was established at Mount Sinai School of Medicine for the purpose of assessing clinical performance with the use of standardized patients, and currently all students in the eight medical schools in the New York City Consortium are tested at The Morchand Center. This paper presents an overview of standardized-patient assessment, a description of the assessment program at The Morchand Center, and a summary of research on this assessment approach including research at The Morchand Center.

Assessment in medical education has expanded its focus in recent years from simply testing cognitive knowledge of medical science to testing actual performance in the clinical situation (1, 2). This performance-based approach is being used increasingly worldwide to assess the clinical skills and competencies of medical students, residents, and physicians, and plans are underway for the use of performance-based assessment in licensure and certification examinations in the United States and Canada (3, 4). Performance-based assessment in medical education is centered around the use of standardized patients (5). A standardized patient (or SP) is a nonphysician who has been trained to portray a specific patient case in a consistent, standardized fashion. In the interaction with the examinee, the standardized patient presents the case history in response to questioning by the examinee and undergoes physical examination at the examinee's direction. Examinees interact with standardized patients the same way they would interact with a real patient. In addition, SPs also complete checklists following the student-SP encounter on which they document actions performed by students on history taking and physical examination and behavior related to interpersonal and communication skills.

At Mount Sinai School of Medicine, The Morchand Center for Clinical Competence was established in 1991 for the purpose of assessing and teaching clinical skills with the use of standardized patients (6). It was established by a one million dollar gift from Mrs. Marietta Morchand in memory of her husband and son and in honor of Dr. Mark H. Swartz. Subsequently, a consortium of eight New York City medical schools was formed, and a fourth-year assessment that uses standardized-patient cases is now administered at The Morchand Center to all students in the eight New York City Consortium schools each year. This paper describes the assessment program at The Morchand Center proceeded by an overview of performance-based assessment that uses standardized-patient cases. In addition, re-

---

From The Morchand Center for Clinical Competence, Mount Sinai School of Medicine (MHS) and Statistics and Research Consulting, Southern Illinois University School of Medicine (JAC). Address reprint requests to Mark H. Swartz, M.D., The Morchand Center, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, NY 10029.

search on the psychometric properties of this testing approach is summarized, including research conducted at The Morchand Center. In particular, a major study of the validity of standardized-patient assessment which is underway at The Morchand Center is described.

Standardized Patients

Standardized patients were first introduced by Barrows and Abrahamson in 1964, in response to ongoing concerns about assessment in medical education (5). Up to the 1950s, assessment relied greatly on judgments by senior physicians in medical schools and on specialty boards, judgments usually rendered on the basis of essay and oral examinations (7). Starting in the 1950s, concerns were raised about the reliability of this approach and, as a result, medical education adopted a two-pronged approach to assessment, using multiple-choice questions (MCQs), to provide an efficient and reliable assessment of knowledge, coupled with faculty ratings of actual clinical performance, to provide a direct measure of clinical competence. Unfortunately, the MCQs were thought to lack validity for measuring the application of knowledge to real clinical problems, and the faculty ratings were not standardized in that different students were rated by different faculty and worse they were rated on different patient problems which might vary in a myriad of different ways. Assessment in medical education at that time could be characterized as concerned with the need for greater psychometric rigor while retaining the realism of the clinical setting, and standardized patients seemed to offer a kind of natural resolution to this assessment dilemma. Standardized patients provide a high-fidelity simulation of the real clinical situation, and they provide a standardized presentation of the same patient problem from examinee to examinee (7).

CPX and OSCE. Two examination formats have evolved that use standardized-patient cases (8). One is the Clinical Practice Examination (CPX) in which the cases simulate the complete doctor-patient encounter like the examinee will face in medical practice. That is, the examinee is given just the patient's complaint and must inquire about history, examine the patient, and order and interpret tests, to arrive at a diagnosis and management plan. Another widely used examination format, the Objective Structured Clinical Examination (OSCE), uses SPs to simply simulate parts of an encounter to test specific tasks or technical skills, such as educating a diabetic patient about insulin or taking a patient's blood pressure (9). The typical OSCE station is shorter than a CPX station (5 or 10 min versus 15 to 20 min), and an OSCE examination may include a station or so which does not involve a standardized patient, such as reading an x-ray.

Surveys of SP Use. Several surveys have shown that standardized patients are being widely used in medical education (10, 11). A survey of the 142 curriculum deans of U.S. and Canadian medical schools conducted in 1989 showed that 94 of the 136 deans responding (70%) indicated that standardized patients were being used in teaching and assessment at their schools (10). A follow-up survey conducted in 1993 showed an increase, with 111 of the 138 schools responding (80%) reporting the use of SPs (11). Many of the schools reported that they even have a central office that coordinates their standardized-patient program.

Licensure and Certification. As mentioned above, a number of medical licensing and specialty certifying boards in the United States and Canada are considering the use of standardized patients (3, 4). Work is underway on test development and implementation by the National Board of Medical Examiners, the Educational Commission of Foreign Medical Graduates, the Medical Council of Canada, the Royal College of Physicians and Surgeons of Canada, the Corporation of Medical Professionals of Quebec, and others. Of particular interest, the National Board of Medical Examiners has established an office for the development and implementation of standardized-patient assessment as a part of licensure, and at their annual meeting in March 1995, the 80-member National Board voted overwhelmingly to endorse the use of standardized-patient cases as a part of USMLE Step II with a time frame of 4 to 7 years as a target for implementation (4).

The Morchand Center Assessment Program

A fourth-year assessment that uses SP-based cases is administered by The Morchand Center at Mount Sinai School of Medicine of the City University of New York to students in member schools of the New York City Center for Clinical Competency (6). This consortium of New York City medical schools is funded by the Josiah Macy, Jr. Foundation and consists of the following eight schools: Albert Einstein College of Medicine, Columbia University School of Medicine, Cornell University Medical School, Mount Sinai

School of Medicine, New York Medical College, New York University School of Medicine, State University of New York—Health Sciences Center at Brooklyn, and State University of New York—Health Sciences Center at Stony Brook. The fourth-year assessment was first administered to Mount Sinai, New York Medical College, and Albert Einstein College of Medicine students in the class of 1991. The examination consisted of eight SP cases, four of which were developed at Mount Sinai, and the other four at the National Board of Medical Examiners (NBME) for their Standardized-Patient Project. Starting with the graduating class of 1995, students in all eight member schools in the consortium were tested, and approximately 1,200 students will be tested in the present year (class of 1996). This represents approximately 10% of all the graduating fourth-year medical students in the United States. Students at some schools are required to take the examination, although their scores do not affect their academic standing; students at the other schools are volunteers. The purpose of the examination is to provide feedback to students on their overall clinical performance and skills, as well as to provide feedback to the faculty concerning areas of weakness in student performance. In the 6 years the examination has been given, students have approached the examination seriously and, from all outward appearances, have performed to the best of their abilities.

The Examination. At present, the examination consists of seven SP cases that represent problems commonly encountered in internal medicine, surgery, pediatrics, gynecology, and psychiatry. Each case requires 30 minutes for its administration, with 20 minutes for the student-SP encounter, during which the student performs a complete focused history and physical examination, and another 10 minutes postencounter, for the students to answer case-related written questions concerning pathophysiology, differential diagnosis, test selection, and test interpretation. During the postencounter period, SPs complete checklists on which they record student actions on history taking and physical examination and note behaviors on interpersonal and communication skills. Cases are selected and developed by faculty representatives from the eight member schools, who meet every 6 weeks to discuss examination direction and policy and who work in *ad hoc* committees to complete the many tasks involved in conducting a testing program of this magnitude. For test administration purposes, students are separated into groups of about 14 students each, and the entire seven-case examination is administered to one group at a time, requiring one day of testing time for each group. At the end of the examination day, students are given videotapes of their performance on the SP cases, which students may review individually or with a faculty advisor. Each student is given one videotape with all seven interactions on it. The intercase reliability of the seven-case examination was .69 for the class of 1995.

The Morchand Center. The Morchand Center is located in the Annenberg Building at Mount Sinai School of Medicine and consists of seven examination rooms and a conference room, each equipped with closed-circuit color television cameras and stereo microphones. All of the audio-visual equipment is linked to a separate Observation Theater which serves to monitor the simulated doctor/patient interactions used to teach important skills. As the student/patient interaction takes place, the proceedings can be watched in the theater, which was custom designed specifically to meet the needs of The Center. The theater's state-of-the-art audio-visual system serves as the eyes and ears of the program. Eight television monitors, with a capability of displaying two different images simultaneously, line one wall. The viewers can adjust the controls on their seats to tune into any of the examination rooms, and headsets are available for listening to various rooms at the same time. In addition, viewers can adjust the cameras in the examination rooms from the Observation Theater to observe the activity from different angles. The Observation Theater has controls for panning, tilting, and zooming silently. Students and faculty can also observe videotapes of their examinations in the theater. The Morchand Center is a powerful display of the best in technology and the latest in medical education.

The actual administration of the SP examination is handled by the staff at The Morchand Center, which has five full-time members. The staff consists of a medical director, an administrator, a standardized patient trainer, a coordinator for the program, and a receptionist/secretary.

Research on Standardized-Patient Assessment

Because of the widespread interest in standardized-patient assessment, a Consensus Conference on the Use of Standardized Patients was sponsored by the Association of American Medical Colleges (AAMC) in December 1992 (1, 2). A major purpose of that Conference was to review the research findings on the psychometric properties of SP assessment, and the consensus of Conference participants was that the findings were "reassuring" and that published studies largely addressed the psychometric uncertainties related to the use of standardized patients in medical education" (1). An overview of research on SP assessment is presented here to give some appreciation of the research issues and the nature of research in this area. For the most part, the research to date can be thought of as falling in three general areas: 1) research on the SPs themselves, 2) examination reliability, and 3) correlations of examination scores with other variables.

The SPs Themselves

First, the central role of SPs in this examination format raises unique psychometric questions about the SPs themselves: can they do the job? That is, are they realistic, are they accurate at portraying their role, and are they accurate at checklist completion? With respect to SP realism, a number of studies have shown that experienced physicians cannot differentiate SPs from real patients, when the SPs are sent unannounced into the physicians' office (e.g., 12, 13). This finding is particularly impressive given that the physicians had agreed in advance to participate in the studies and thus knew that SPs would be coming into their offices, unannounced, sometime the next few months. Other studies have looked at performance of house officers with real patients compared with matched SP clones of the real patients, and results showed performance on history taking, physical examination, findings, and diagnosis was quite similar with the real and the standardized patients (e.g., 14).

With respect to accuracy of case portrayal, one study based on a major doctoral dissertation by Tamblyn showed that on average SPs were more than 90% accurate in portraying details of the case (15, 16). More than half had average accuracies between 95% and 100%. With respect to checklist completion, another study based on Tamblyn's dissertation showed that for SPs trained at the same site, the average agreement in recording examinee actions was 82%, and similar results were found for SPs trained at different sites (15, 17). Moreover, the results of another study showed that SPs with limited training were just as accurate as physicians in evaluating and recording examinees' clinical performance on checklists (18). All in all, studies in this area show that SPs can do their job: they are realistic, and they are accurate at case portrayal and checklist completion. Moreover, informal reports by standardized-patient trainers suggest that any shortcomings in this area can be overcome by further training.

Reliability

Reliability is probably the most researched and most documented psychometric property of standardized-patient examinations (19–21). Reliability refers to whether measurements obtained with a given measurement process are reproducible with repeated application of the process. That is, it refers to the reproducibility, consistency, or stability of measurements obtained with repeated measurement. The reliability of a standardized-patient examination is typically evaluated with a generalizability coefficient, which can be thought of as a Pearson Correlation Coefficient between scores on the examination at hand *and* scores on a comparable examination that consists of a different sample of SP cases *but* from the same population of possible cases (21).

The review of SP research by van der Vleuten and Swanson in 1990 reported reliabilities obtained for 13 SP examinations at nine institutions (20, 21). The reliabilities (i.e., generalizability coefficients) ranged from .41 for the SP examination at the University of Texas Medical Branch to .85 for the SP examination administered at the University of Toronto. Reliabilities of .80 or above are typically recommended for educational and psychologic tests. These reliabilities fall a little short of the mark, but they are based on the earliest examinations administered. Our sense is that reliabilities of more recent examinations are much higher.

Also, the reliabilities here are *score* reliabilities; that is they indicate the reproducibility of the actual quantitative, numerical scores on the respective SP examinations. In actual practice, the reliability of the scores on an SP examination is probably of less concern than the reliability of the pass-fail decisions (19, 21). That is, the accurate classification of examinees as those who have passed or failed is typically of much greater concern than the accurate relative ordering of examinees as indicated by their examination scores. And, fortunately, the reliability of pass-fail decisions is typically much better than the reliability of examination scores. For example, using data on total scores for the seven-case examination administered at The Morchand Center to students in the eight schools in the New York City Consortium in the class of 1995, the reliability of the pass-fail decisions with passing standard at one and one-half standard deviations below the mean was .96, whereas the reliability of the examination scores as evaluated with the generalizability coefficient was only .69. Clearly, the pass-fail decision reliability of .96 achieves the recommended reliability level of .80, although the score reliability of .69 does not.

Reliability and Number of SP Cases. An important factor that affects the reliability of SP examinations (both score reliability and pass-fail reliability) is the number of cases in the examination. Not surprisingly, an examination consisting of only one case just won't work. A single score and a single pass-fail decision based on only one SP case do *not* provide a reliable assessment of the entire universe of an examinee's clinical skills and competence. To achieve the recommended .80 reliability, a number of SP cases are needed. For example, again using total scores from The Morchand Center, 13 SP cases are needed for the score reliability to reach .80, but only two cases are needed for a pass-fail reliability of .80. It is important to note that just as an examination's pass-fail reliability is typically higher than its score reliability, the number of cases needed to achieve the recommended .80 reliability is typically less for pass-fail reliability than for score reliability.

The results suggest that the pass-fail reliabilities for most examinations can reach the recommended .80 level with an examination of reasonable length, that is, without requiring a prohibitive number of SP cases. Moreover, the accuracy of pass-fail decisions would seem to be our primary, and maybe even our only, concern. However, at this stage in the development of SP testing, many examinations do *not* have a solid pass-fail cutoff that is established with rigorous standard-setting procedures. Clearly, this is an important area that needs more work by test developers. The establishment of solid passing standards may be the key to achieving acceptable levels of reliability with examinations of reasonable length.

Correlations with Other Variables

The third area of research concerns studies that simply looked at relationships between SP-examination scores and other variables. The importance of this research though is that it lets us see what is (and is not) being assessed by SP examinations. Or it lets us test whether the SP-examination scores are consistent with our thinking about clinical competence, by testing whether the scores are related to what they should be related to and not related to what they shouldn't be. Although many issues fall in this area, only a few of the major issues are considered here.

Gender. An issue of concern that SP examinations share with other medical examinations is the effect of gender. That is, is there a difference in the performance levels of male and female examinees? But with SP examinations, the issue is more complicated, because the gender of the standardized patient must be considered as well. The concern is that examinee gender and SP gender might interact, such that the performance of examinees of a given gender might be higher (or lower) depending on the gender of the SP. For example, male examinees might perform better than females on cases simulated by male SPs, and female examinees might perform better on cases simulated by female SPs. But research has shown no evidence of such interactions (22–24). These findings are important because it means that examinees of a given gender will not be favored as a result of the mix of male and female SP cases. That is, if male examinees performed better on male SP cases, an examination with more male SP cases would favor male examinees, and likewise for females. The studies present no evidence to suggest that this is a problem. These studies also reported no clear evidence of a main effect of examinee gender: there was no clear evidence that male examinees did better than females or that females did better than males. However, several of the studies reported weak evidence that suggested that female examinees appeared to receive higher ratings than males on interpersonal and communication skills (22, 24), but the findings are weak at best. Overall, the studies showed that male and female examinees generally performed equally well, and that they performed equally well regardless of the gender of the SP.

Test Security. Another concern has been test security. SP examinations are typically lengthy, requiring the administration of the same SP cases to several consecutively tested groups of examinees. For example, for students at Mount Sinai, each class of about 140 students is randomly divided into 10 groups of about 14 students each, and the examination in its entirety is administered to one group at a time, requiring 10 days of testing time for an entire class. This schedule for test administration has raised concerns that students in groups tested early in the examination period might pass on information about the cases to students in groups tested later, in which case, students in the later groups would have an advantage over the students in the earlier groups. Widespread violations of test security like this would constitute a serious threat to the validity of the scores. Fortunately, the test security studies conducted to date have shown no evidence of such widespread violations of test security (25–28). Students in groups tested later do not perform better (or worse) than students in groups tested earlier. A survey of students completing an SP examination, though, showed that there is a "grapevine effect": students do receive information about the examination from other students. Nevertheless, scores did not improve throughout the examination period (25). Similar informal reports have been obtained from the other sites at which test security studies were conducted: students do talk about the examination, but with no systematic improvement in scores throughout the examination period.

A test security study conducted at Mount Sinai even went so far as to arrange for students who had been tested on one day to meet with students who would be tested on the next day and to communicate any information about the SP cases that they thought would be helpful (27, 28). Despite this deliberate, somewhat massive assault on test security, performance of students in the second group was unaffected. The authors concluded that SP examinations by their very nature may be impervious to such violations of test security, given that the SP cases require a complicated, integrated response by the examinees which may not be facilitated by prior information about the cases as communicated by other examinees. Test security, then, does not seem to be a problem. However, because the stakes will increase with licensure and certification testing, we must continue to carefully monitor these results.

Multiple SPs. One of the practical concerns about SP examinations stems from the common practice of using two or more standardized patients (i.e., multiple SPs) to simulate the same case (29, 30). Because of the time required to administer an SP examination such as to a class of several hundred medical students, two or more (multiple) SPs are generally used to portray a given case. Naturally, this raises questions about the consistency of the simulation from one SP to another. More fundamentally, the concern is about the impact of multiple SPs on examination reliability. That is, is the reliability of examinees' total scores affected by the use of multiple SPs on individual cases? Fortunately, studies have shown little or no effect of multiple SPs on examination reliability. One study, for example, found no difference between the reliability of cases for which only a single SP was used versus the reliability of other cases for which multiple SPs were used (29). A second study found, for a given set of cases, no difference in reliabilities when the cases were simulated by single SPs versus when the same cases were simulated by multiple SPs (29). Several other studies used a statistical approach applied to cases simulated just by multiple SPs to estimate directly the potential effect of multiple SPs (29, 30). Like the other two types of studies, these studies also showed little or no effect of using multiple SPs to simulate the same case on examination reliability. For example, in one of these studies, for total scores the generalizability coefficient corrected for the use of multiple SPs was .65, and the generalizability coefficient uncorrected for multiple SPs was .64. In brief, the use of multiple SPs on the same case seems to have little if any effect on examination reliability.

Given that multiple SPs appear to create no problem for reliability, what about having a single SP simulate a given case? The critical issue here is that each SP must simulate his or her case for every examinee, which could be as many as 50 to 100 to even 200 medical students in a given class. Repeated simulations of the same case by a given SP naturally raises questions about the ability of the SP to portray the case and to complete checklists consistently throughout the entire examination period. It would seem reasonable to speculate that, with repeated simulations, the SPs might grow tired and become careless in portraying their cases and inconsistent in completing checklists, so that measurement error would increase throughout the examination. On the other hand, it could be argued that the practice and experience gained with repeated simulations could serve to sharpen the SPs, resulting in more consistent simulations and checklist completions, so that measurement error would decrease instead. This issue was examined in one study that found no clear evidence of an effect of repeated simulations on reliability (31). In brief, the use of multiple SPs or single SPs throughout the entire examination appears to have little or no effect on reliability.

Sequential Testing. An approach that has been suggested for dealing with the concerns about the length of SP examinations is sequential testing (32, 33). Sequential testing refers to an approach in which all examinees take a short screening test, which might consist of the first few cases in the full examination. Examinees who scored above a predetermined cutpoint on the screening test would immediately receive a pass on the full examination and would be excused from taking the remaining SP cases in the examination. Examinees who scored below the screening test cutpoint would continue the examination and complete all of the SP cases. Pass-fail decisions for these examinees would be based on their full examination performance. The feasibility of sequential testing with an SP-based examination was demonstrated in a recent study that involved a retrospective analysis of data from an SP-based examination, for six classes of senior medical students (32). The results showed that for all six classes, 228 of the 404 students tested (i.e., 56% of all students) would have passed the six-case screening test and would have been excused from the rest of the 18 cases in the full examination. Of these 228 students, 227 or 99.6% went on to pass the full examination. Thus, only 1 of the 228 students passing the screen failed the full 3-day examination, resulting in a passing error rate for the screen of only 0.4%. Using these results, it has been estimated that for an 18-case examination, for a class of 170 students, sequential testing could reduce costs by more than $20,000. Clearly, the results have important practical implications, and the NBME is considering the use of sequential testing for its SP licensure assessment.

Validity. As mentioned above, participants at the AAMC's Consensus Conference on The Use of Standardized Patients concluded that the findings were "reassuring, and that published studies largely addressed the psychometric uncertainties related to the use of standardized patients in medical education" (1,2). But, despite this positive evaluation, research on SP testing has not addressed the ultimate, bottom-line question of validity, and research has not informed case and examination development, in particular scoring and standard setting, because it has produced little or no information on how to improve cases and examinations by making them more valid (34). Validity refers to whether an examination measures what it purports to measure, and the property or characteristic the examination claims to measure is called a construct. The validity of an SP examination, then, refers to whether the examination really measures the construct of clinical competence, which would be demonstrated—and this is the important point—by evidence showing that variation in SP-test scores is indicative of actual variation in clinical competence. Thus, a measure that closely approximates the clinical competence construct is needed for use as a gold-standard criterion to provide convincing evidence of validity of SP examinations. One recurring suggestion for the gold-standard criterion has been global ratings by faculty-physician observers, which may be the only way that we will be able to validate SP examinations (34).

The first major validity study of SP assessment using faculty ratings as the gold-standard criterion currently is being conducted at The Morchand Center (34, 35). The study is based on data from the seven-case SP examination that is administered to senior medical students in the eight medical schools in the New York City consortium. Videotapes of 50 students randomly selected from one of the eight schools will be viewed and rated by five faculty physicians, each from one of the seven remaining schools. Faculty will rate students on each case and overall; and SP-examination scores will be obtained as they usually are, by using the scoring system developed by faculty representatives of the eight consortium medical schools. Students' scores and passing decisions will be related to faculty global ratings to check whether current scoring and standard setting in actual use agree with faculty observation of clinical performance.

Now, if this doesn't work, the other side of the coin of validity studies like these is that they provide data for examination development and refinement, including scoring and standard setting (34). Faculty ratings could be used to perform an item-validity-type analysis, to identify those items that best capture the gold-standard criterion. Not only would this allow us to construct SP cases using only items with empirically demonstrated validity, this approach could also be used to study the general nature of items that most enhance validity and those that most detract from validity, which could increase our sophistication in the construction of new cases. Similarly, faculty observers could be interviewed in depth soon after the examination about the basis of their ratings, in particular, the failing and borderline ratings. The interviews would seek to determine the critical factors that account for the ratings, which could serve as a basis for developing checklist items and postencounter exercises. Thus, with faculty observing and rating performance, we would be in a position to establish empirically based scoring and passing standards that reflect actual clinical performance, rather than trying to guess in advance, as we do now, the characteristics that indicate good performance.

Using faculty ratings as the gold standard would also provide an interpretation of SP-examination scores, namely, the scores could be seen as a proxy for the gold-standard faculty ratings. Case and examination scores could be reported in terms of what might be called "faculty-rating equivalents," such as "high pass," "pass," "low pass," and "fail," or just "pass" and "fail" rather than in terms of the case or examination rawscore scale (36). The meaning of scoring and standard setting with this approach would seem to be exactly what is needed for SP testing; namely, scores and standards that reflect physician ratings of real students under controlled, standardized conditions.

Conclusion

Standardized-patient assessment is being used increasingly worldwide, and the research base that accompanies this usage is reassuring, showing that SPs are realistic and are accurate at portraying the patient role and completing checklists, that the examinations can achieve acceptable levels of reliability provided they are of sufficient length, and that relationships with other measures tend to be consistent with our thinking about clinical competence. At least, the research to date shows that SP testing has not succumbed to some of the major pitfalls that many researchers initially thought would undermine this assessment approach. Also, research on the ultimate, bottom-line question of validity is getting underway, and even more important, this research is using a format that should enhance our understanding of validity and contribute to the development and refinement of SP assessment.

References

1. Proceedings of the AAMC's consensus conference on the use of standardized patients in the teaching and evaluation of clinical skills [Special Issue]. Acad Med 1993; 68:437–483.
2. Annex to the proceedings of the AAMC consensus conference on the use of standardized patients in the teaching and evaluation of clinical skills [Special Section]. Teach Learn Med Int J 1994; 6:2–35.
3. Highlights of the 1995 Annual Meeting of the Board. Natl Board Exam 1995;42:1–3.
4. Reznick RK, Blackmore D, Dauphinée WD, Rothman AI, Smee S. Large scale high stakes testing with an OSCE: Report from the Medical Council of Canada. Acad Med 1996; 71:S19–S21.
5. Barrows HS, Abrahamson S. The programmed patient: A technique for appraising student performance in clinical neurology. J Med Educ 1964; 39:802–805.
6. Swartz MH, Friedman M, Gambert S, Anderson-Kline S, Regan MB, Stimmel B. Assessment of clinical competence utilizing standardized patients: What are we really measuring? In: Harden RM, Hart IR, Mulholland H, eds. Approaches to the assessment of clinical competence. Norwich, Great Britain: Page Brothers, 1992: 261–263.
7. Mast TA, Schermerhorn GS, Colliver JA. Medical education. In: Alkin MC, ed. Encyclopedia of educational research, 6th ed., New York: Macmillan Publishing Co., 1992.
8. Barrows HS. Multiple stations. In: Hart IR, Harden RM, ed. Further developments in assessing clinical competence. Montreal: Can Heal Publications, 1987.
9. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). Med Educ 1979; 13:41–54.
10. Stillman PL, Regan MB, Philbin M, Haley HL. Results of a survey on the use of standardized patients to teach and evaluate clinical skills. Acad Med 1990; 65:288–292.
11. Anderson MB, Stillman PL, Wang Y. Growing use of standardized patients in teaching and evaluation in medical education. Teach Learn Med 1994; 6:15–22.
12. Norman G, Muzzin L, Williams R, Swanson D. Simulation in health sciences education. J Instruct Dev 1985; 8: 11–17.
13. Norman GR, Neufeld VR, Walsh A, Woodward CA, McConvey GA. Measuring physicians' performance by using simulated patients. J Med Educ 1985; 60:925–934.
14. Norman GR, Tugwell P, Feightner JW. A comparison of resident performance on real and simulated patients. J Med Educ 1982; 57:708–715.
15. Tamblyn R. The use of standardized patients in the evaluation of clinical competence: The evaluation of selected measurement properties [dissertation]. Montreal: McGill University, 1989.
16. Tamblyn RM, Klass DJ, Schnabl GK, Kopelow ML. The accuracy of standardized-patient presentations. Med Educ 1991; 25:100–109.
17. Tamblyn RM, Klass DJ, Schnabl GK, Kopelow ML. Sources of unreliability and bias in standardized-patient rating. Teach Learn Med 1991; 3:74–85.
18. Elliot DL, Hickam DH. Evaluation of physical examination skills: Reliability of faculty observers and patient instructors. JAMA 1987; 258:3405–3408.
19. Colliver JA, Verhulst SJ, Williams RG, Norcini JJ. Reliability of performance on standardized-patient cases: A comparison of consistency measures based on generalizability theory. Teach Learn Med 1989; 1:31–37.
20. van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: State of the art. Teach Learn Med 1990; 2:58–76.
21. Colliver JA, Williams RW. Technical issues: Test application. Acad Med 1993; 68:454–460.
22. Stillman PL, Regan MB, Swanson DB, Haley HA. Gender differences in clinical skills as measured by an examination using standardized patients. In: Hart I, Harden RM, Des Marchais J eds. Current developments in assessing clinical competence. Montreal: Can-Heal, 1992; 390:395.
23. Colliver JA, Marcy ML, Travis TA, Robbs RS. The interaction of student gender and standardized-patient gender on a performance-based examination of clinical competence. Acad Med 1991; 66:31–33.
24. Colliver JA, Vu NV, Marcy ML, Travis TA, Robbs RS. The effects of examinee and standardized-patient gender and their interaction on standardized-patient ratings of interpersonal and communication skills. Acad Med 1993; 68:153–157.
25. Williams RG, Lloyd JS, Simonton DK. Sources of OSCE examination information and perceived helpfulness: A study of the grapevine. In: Hart I, Harden RM, Des Marchais J, eds. Current Developments in Assessing Clinical Competence. Montreal: Can-Heal, 1992; 363–370.
26. Colliver JA, Travis TA, Robbs RS, Barnhart AJ, Shirar LE, Vu NV. Test security in standardized-patient examinations: Analysis with scores on working diagnosis and final diagnosis. Acad Med 1992; 67:7–9.
27. Swartz MH, Colliver JA, Cohen DJ, Barrows HS. The ef-

fect of deliberate, excessive violations of test security on a standardized-patient-based performance examination. Acad Med 1993; 68:76–78.
28. Swartz MH, Colliver JA, Cohen DS, Barrows HS. The effect of deliberate, excessive violations of test security on a standardized-patient examination: An extended analysis. In: Rothman A, Cohen R, eds. Proceedings of the Sixth Ottawa Conference on Medical Education. Toronto: University of Toronto Bookstore Custom Publishing, 1995:280–284.
29. Colliver JA, Morrison LJ, Markwell SJ, et al. Three studies of the effect of multiple standardized patients on intercase reliability of five standardized-patient examinations. Teach Learn Med 1990; 2:237–224.
30. Colliver JA, Marcy ML, Vu NV, Steward DE, Robbs RS. The effect of using multiple standardized patients to rate interpersonal and communications skills on the same case on the intercase reliability of the ratings. Teach Learn Med 1994; 6:45–48.
31. Colliver JA, Steward DE, Markwell SJ, Marcy ML. Effect of repeated simulations by standardized patients on intercase reliability. Teach Learn Med 1991; 3:15–19.
32. Colliver JA, Mast TA, Vu NV, Barrows HS. Sequential testing with a performance-based examination using standardized patients. Acad Med 1991; 64–66.
33. Colliver JA, Vu NV, Barrows HS. Screening test length for sequential testing with a standardized-patient examination: A receiver operating characteristic (ROC) analysis. Acad Med 1992; 67:592–595.
34. Colliver JA. Validation of standardized-patient assessment: A meaning for clinical competence. Acad Med 1995; 70:1062–1064.
35. Swartz MH, Colliver JA. A proposal to assess the validity of standardized-patient examinations. Funded by Astra/Merck Pharmaceuticals, August 1994.
36. Colliver JA, Barnhart AJ, Marcy ML, Verhulst SJ. Using a receiver operating characteristic (ROC) analysis to set passing standards for a standardized-patient examination of clinical competence. Acad Med 1994; 69:S37–S39.

Breast Cancer Chemotherapy after Fifty Years:

More Cures and Less Mutilation

Ezra M. Greenspan, M.D.

Breast cancer is prognostically the most heterogeneous common cancer encountered in the western world. The importance of host systemic influences on breast cancer were first demonstrated in 1896 by Beatson's oophorectomy (1), which transiently inhibited 20% of advanced cancers, yet the influence of host characteristics on prognosis has received relatively little research attention in this century despite their obvious importance. The classical surgical locoregional therapies based on Halstead's mastectomy and recent modifications by lumpectomy and radiation, yield cure rates varying less than 5% from each other when analyzed by staging stratifications (2, 3). The earliest hormone therapies initiated in 1942 by the Massachusetts General Hospital with estrogen and later testosterone had minimal effects on control in advanced cancers. Before 1960, before the era of chemotherapy, primary surgical procedures and radiotherapy in stages I, II, and III breast cancer were yielding a 45% 5-year survival. In those years the 5-year survival was considered a landmark because distinct flattening and a change in direction of the survival curve was noted at the 5-year mark. Statisticians had already pointed out (4) that this survival curve in breast cancer showed a two-phase characteristic suggestive of two major categories of risk, namely high-risk, and low-risk patients. Now in the era of combination chemotherapy, survival after 10 years probably should be the important prognostic landmark for both the layman and informed clinician. With increasingly large numbers of patients being diagnosed in stage I (70%), at least 75% of newly diagnosed patients are expected to survive 10 years. Thus, the 10-year survival is important but its clinical interpretation must depend on the category of risk when treatment is begun.

Originally, tumor size, age, and lymph node status with pathologic features (5) were the hallmarks of high- versus low-risk classifications (6). Numerous biochemical characteristics and more recently hormone receptors and enzymatic and genetic analysis have yielded valuable predictors of subgroups of patients with breast cancer who are more likely to die, but usually within the first 5-year period. Despite hundreds of retrospective and prospective studies, the major significance of age, tumor size, and lymph node status remain paramount primarily for the near-term prognosis. None of these prognostic factors reliably predict long-term survival (7).

The regression response itself of breast cancer to combination chemotherapy, as a prognostic factor or marker, only recently has come into the limelight, despite the fact that it was originally reported in 1960 to 1963 (8) at Mount Sinai and in my office practice to show that regressions as high as 75% in advanced breast cancer were achievable with combinations of thiotepa plus methotrexate. To obtain this high rate of objective response it was necessary simply to cycle oral methotrexate tablets on several days per month to minimal but definite tolerable stomatitis which was always reversible. My first presentation on 96 patients was greeted with ridicule, dis-

---

Clinical Professor of Medicine (Oncology), The Mount Sinai School of Medicine, New York, NY; Medical Director, The Chemotherapy Foundation. Address reprint requests to Ezra M. Greenspan, M.D., Medical Oncology, Mount Sinai Medical Center, Box 1178, One Gustave L. Levy Place, New York, NY 10029.

belief, and derision. In fact at the AACR meeting in 1960, I was publicly called "a liar." The action of fluorouracil and Cytoxan in three and four drug combinations on liver metastases and bone metastases was next reported (9, 10). Soon, Richard Cooper in Buffalo, also in private practice, reported in 1965 that his response rate was 80% in advanced cancers treated with the five-drug combination of CMFVP (11). He always pushed vincristine until loss of reflexes occurred. Nevertheless, any major long-term prognostic importance of the chemotherapy response itself seemed specious when applied to advanced breast cancer because almost all patients, even good responders with virtually complete or good regressions, relapsed after several years and less than 10% survived disease free after 5 years. However, some of the skepticism regarding the future value of chemotherapy was next overcome by a different timing approach. Cooper was so well respected by his surgical colleagues in Buffalo that he was able to assemble a substantial series of postsurgical patients with stage IIB breast cancer with four or more positive lymph nodes. He showed in the early seventies that the 3- to 5-year disease-free survival with adjuvant CMFVP was markedly improved in stage IIB from the historical 35 to 40% level expected to a 75% level (12). Patients were usually started on chemotherapy within 3 weeks after operation. But when there was a delay for radiotherapy, and chemotherapy was begun a month later, they had a shorter survival time.

Meanwhile, perioperative single-drug studies with thiotepa and Cytoxan had been initiated in the early sixties by Fisher et al. (13) and Nissen-Mayer et al. (14). But it was determined only 10 to 20 years later that the survival benefit was less than 12%. The most important feature of the CMFVP survival curve was that it was sustained, hardly dropping between the years 5 and 10. Cooper's major advance set the stage for Nixon's War on Cancer in the mid-70s. The modern era of large-scale surgical adjuvant chemotherapy trials was then launched with much fanfare. Most important were the studies, such as those from the Bonadonna group, which showed little decline in survival in large randomized trials of CMF in the 5- to 10-year periods after treatment (15). Although breast cancer chemotherapy response rates were now proven to be in the range of lymphomas and Hodgkin's disease, the idea of designating breast cancer as a curable disease was still not accepted and even seemed far fetched because it was impossible to predict which patients were going to be the good responders beyond the 45% expected to be cured by locoregional methods. It was in the small pathologic and clinical syndrome of inflammatory breast cancer representing less than 5% of all breast cancers that the true curative potential of combination chemotherapy was observed most dramatically. In this relatively rare subgroup of patients with stage III breast cancer, who formerly had a 95% mortality rate in 1 year, CMFV combination chemotherapy induced an almost 100% response and the responses were sustained particularly after adriamycin was introduced in 1973 (16). Survival was markedly extended to the point that current inflammatory cancer studies show disease-free survival in at least 40 to 50% of patients 10 years after treatment. Certainly this seemed to advance the concept of cure. The introduction of tamoxifen after 1974 added immeasurably to long-term survival, perhaps increasing the net response to chemohormonal therapy by as much as 20% to 30% among stage II cancers (17).

The substantial subgroup of patients in stage IIIA and stage IIIB also began to show remarkable longevity especially after adriamycin-containing polychemotherapy frequently given in a sustained manner for at least 1 year. Although these patients with locally advanced disease were presumed to have poor immunocompetence, contrary to expectations, they were noted in several studies (18–20) to show better delayed cutaneous immunocompetence than patients with stage II cancer. This unusual fact was generally ignored by oncologists and the role of immunity remained in obscurity despite some relevant observations by myself and others (21, 22).

The concept that a dramatic response to combination chemotherapy was a good prognostic sign could be most accurately tested in the stage III setting or whenever the first sequence of systemic management was a neoadjuvant combination of chemotherapy against objectively measurable large primary tumors. Gerson Lesnick and I first observed several unusual clinical responses in stage III in the late fifties when thiotepa and methotrexate was sequenced with Cytoxan and fluorouracil (8). Our sickest patient survived 7 years and died just before the advent of adriamycin. Similar cases spurred a multimodality study involving neoadjuvant therapy of stage III breast cancer by the Eastern Cooperative Oncology Group using radiation or surgery postchemotherapy in the late 1970s. But routine neoadjuvant therapy for early-stage breast cancer, such as stage II T3 or large stage III lesions, was never seriously considered in the United States until 10 years ago. Hortobagyi (23) from the M.D. Anderson Cancer Center has reported a surprising 30 to 60% 10-year disease-free survival in stage IIIA or stage III breast cancer (Fig. 1).

However, routine neoadjuvant chemotherapy was adopted earlier in France (24) and in Italy where mastectomy was viewed frequently as breast mutilation and there was a greater public emphasis on breast preservation. Neoadjuvant chemotherapy soon spread to Canada. This approach is now being used for the first time on a large scale in England (25) and to a lesser scale in the United States. Many patients with large early cancers (stage II or III) are being treated with primary chemotherapy. The data appear to show that the needed favorable longevity characteristic is a complete response to neoadjuvant chemotherapy confirmed pathologically with multiple aspiration biopsies. This cytoreductive response can be used as a predictive indicator of long-term disease-free survival. The most recent dramatic illustration of this approach was reported in 1995 (26) from the Royal Marsden group in London. They found that better than 95% major regressions of advanced localized cancers occurred (T3–T4 lesions, stages II and III). All of these patients were in the category of "early large cancers" yet showed major responses to chemotherapy of an intensive novel three-drug regimen conprising a combination of 6 months of a steady, slow infusion of fluorouracil by pump accompanied every 3 weeks with bolus epirubicin and cisplatinum (ECF). These unusual results were superior because FAC did not approach the 66% complete pathologic response rate found by the British with ECF. FAC yielded only 30 to 40% responses in several studies and the CMFV Cooper-type regimen yielded responses in the 16% range. Less than 8% of the British patients required mastectomy after radiotherapy. Thus it should be noted that the most conservative and formerly antichemotherapy group in the Western world, namely the British, have now turned full circle and are enthusiastic about early and primary combination chemotherapy. A controlled study of ECF versus FAC is now underway in several centers.

The new trend toward protoadjuvant neoadjuvant polychemotherapy should be studied and hopefully reinforced with immunotherapy which could be best evaluated if it were used before radiotherapy and anesthesia. However such ideal timing is not a realistic possibility. It should be noted that experimental studies indicate that significant immunosuppression occurs associated with operation and with radiotherapy. Immunotherapy appears to be a contributor to the long-term survival of the patients with inflammatory breast cancer, especially in Japanese and French studies which usually use BCG. The addition of hormone therapy with tamoxifen confirmed by the recent large meta-analyses also facilitated long-term outcomes in breast cancer. The concept that tamoxifen is primarily an apoptotic agent may be the key to its beneficial effects which now extend well over a 10-year period with a risk reduction of at least 30% in diverse patients with breast cancer. Immunologic staging of patients with breast cancer preoperatively has never been performed as an established practice by oncologists. In fact oncologists generally ignore immunotherapy despite numerous experimental studies which indicate great promise in such therapeutic concepts, especially if optimum timing of immunotherapy can be achieved. In experimental and in certain clinical settings there is little doubt that the timing of immunotherapy and the route of immunotherapy are critical. Overkill (overdosage) (27) with resulting immune

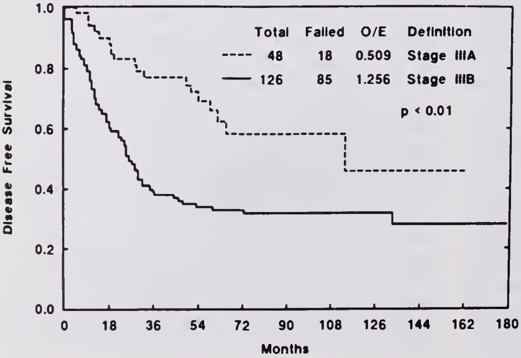


Fig. 1. Fifteen-year projected long-term survival after aggressive FAC therapy in Stage III breast cancer. Clinical 10-year "cure" has been achieved in 30 to 45% of patients. Reprinted with permission. Hortobagyi, Am J Clin Oncol 1989; 12(2):123–128.

blockade can be common when chemotherapists use immunotherapy in persistent, repeated multiple doses with such agents as MER or BCG, as done by some of the cooperative groups. Nevertheless, trials of BCG in breast cancer have been positive in some drug protocols not involving such powerful immunosuppressants as the platinum derivatives. In controlled trials in ovarian cancer, immunotherapy with BCG has a proven survival benefit when adriamycin/Cytoxan therapy is used, but not simultaneously with platinum therapy, probably because of the extreme and sustained immunosuppression induced by the platinum. Immunotherapy in the second year after cessation of chemotherapy has never been used on a methodical scale.

Numerous anecdotal cases of unexpected long-term survival have been found in patients with breast cancer (21, 28) who show Ghon foci on routine chest x-ray, a past history of tuberculosis, and a positive tuberculin test. The more slowly growing, benign, and relatively low rate of breast cancer in the Japanese population (29) must be related to many factors, nutritional and otherwise, but certainly they include the exposure to tuberculosis because 97% of the Japanese population had been exposed during and after World War II. Breast cancer in Great Britain may be somewhat more slower growing and perhaps even more readily responsive, partly because of previous exposure to infectious agents including tuberculosis. Most of the British patients received at least one dose of BCG during or after World War II, which could perhaps explain why the average age of breast cancer development in England is 59 years compared with 56 years in the United States.

Individuals who have developed erysipelas after mastectomy, and especially those who have repeatedly developed erysipelas in lymph edematous operated arms, have been reported to survive for 15 to 20 years despite the presence of innumerable involved nodes at the original operation (30). Diabetes and obesity predispose to recurrent infection and perhaps, thereby, this unusual self-immunization.

The public is not aware of the markedly improved mortality statistics recently reported to have occurred since 1990. The media and some advocacy groups for patients with breast cancer have ignored these important data and unceasingly complained that the age incidence rate in America in the past two decades and the overall mortality rates for breast cancer did not show any significant change. However, a startling decline in overall breast cancer mortality rates has now been reported by the Oxford group's epidemiologic and statistical study from Wales and England (31). Since 1990, the mortality rates have dropped successively year by year, falling a total of about 20% (31) (Fig. 2). Somewhat larger reductions were in younger women, ages 45 to 60 years, compared with those older than 60 years. Certainly this is an effect of earlier diagnosis and better treatment being applied in the 1980s, particularly more aggressive chemotherapy and earlier chemotherapy and the use of tamoxifen. The British patients may represent an unusually disciplined population. Hopefully we shall see a similar dramatic drop in mortality rates in the United States when the data are analyzed.

Mortality statistics must always be viewed with caution because much nonoptimal treatment still occurs in America even today. Aside from the influence of delay in treatment and use of alternative treatments prevalent in the less disciplined American public, the variations in clinical practice by different oncologists must be considered. In 1986 the Chemotherapy Foundation informally surveyed the clinical chemotherapy practices of a substantial number of responding oncologists. More than half failed to use adriamycin in high-risk patients when it was indicated, even though Bonadonna and many U.S. cooperative groups had emphasized to the profession, in the eighties, the value of aggressive, early use of adriamycin as first-line therapy in preference to CMF. Even the subtle influence of the clinical workload can alter the influence of patterns of treatment on survival benefits. There was considerable variation in survival between surgeons with the better 5-year survival outcomes among

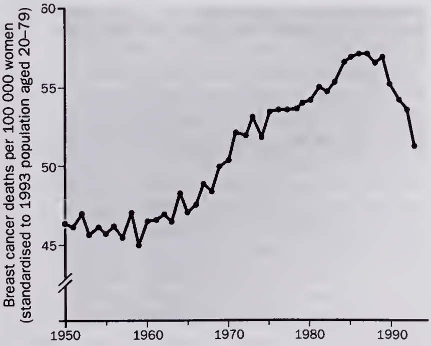


Fig. 2. Annual breast cancer mortality rate as determined by the Oxford University Statistical Group in England and Wales, 1950–1993. A steady decline since 1990 is observed. From Beral V. Lancet. 1995; 345:1642–1643.

those who had a case load of more than 30 patients per year. In an analysis of more than 12,000 patients in Great Britain (32) treated between 1979 and 1988, patients of surgeons with high ratios of chemotherapy and hormone therapy usage showed a doubling of survival from 4.6% to 9.3%. The mean chemotherapy usage ranged from 0 to 46%. Those with a lesser case load had poorer results. This was deemed to reflect the fact that the full range of treatment options and the multimodality components were not available or used by surgeons with smaller numbers of patients.

It should not really be surprising to find that it took at least 10 years of aggressive chemotherapy and hormonal therapy from 1980 to 1990 before the absolute drop in breast cancer mortality began to appear in the epidemiologic studies in Great Britain. The fact that the drop continued steadily year after year is most encouraging (33).

Ultimately we come to the definition of cure (34, 35). Is it virtual cure or real cure? This definition may have to be redefined in view of the current concepts involving apoptosis of cancer and tumor dormancy which are relatively new factors encouraged by the widespread use of biologic response modifiers, especially tamoxifen, angiogenesis inhibitors, and other agents. We do not fully know the significance of carcinoma *in situ* or low-grade intraductal carcinoma. It is known that apparently 25% of Danish women 65 years of age, in routine autopsies, showed microscopic evidence of (usually) low-grade breast cancers (36). Most of these cancers are apparently at rest or in a state of plateau-apoptosis. After an accidental death, patients cured of Hodgkin's disease have been found at autopsy to show microscopic residual disease. So the total disappearance of any pathologic evidence of cancer may not be the requirement for a definition of a cancer cure. Certainly tuberculosis is greater than 95% curable with combination chemotherapy, but old scars and fibrotic tissues may remain for many years, sometimes even showing evidence of bacilli. Thus it has been debated (35) whether the term cure as commonly used may be virtual cure and clinical cure without true pathologic cure. In any event it seems established that careful perusal of the primary tumor pathology 4 months after ECF chemotherapy in the neoadjuvant setting may make surgery superfluous for 92% of large early cancers, especially after radiotherapy has been administered. Moreover, group studies of radiation therapy now show a reproducible impact on the disease-free interval in low-grade cancers, apparently in the 10 to 15% range, a much more favorable view of the survival effects of radiotherapy than in former years (37).

The neoadjuvant pathologic and clinical response could well become the designation-marker of which individual patients will be truly "cured." The noteworthy contribution of these new studies on neoadjuvant therapy is to emphasize the exquisite sensitivity of breast cancer to chemotherapy. Thus the neoadjuvant response to chemotherapy could become a prognostic factor of major long-term importance, especially if the elimination of gross clinical tumor is also associated with host factors that enhance systemic resistance to cancer in future years.

Instead of an annual mortality of 46,000 patients among the 185,000 patients developing breast cancer in the United States in 1996, we should soon be able to reduce the mortality by at least 15,000 more patients compared with the 3,000 to 4,000 being saved today with the methods of past decades. Certainly if chemotherapy and immunotherapy response criteria indicate a good prognosis, less than 10% of patients among the 185,000 new patients every year would eventually die of breast cancer in the foreseeable future.

Glossary

AACR = American Association of Cancer Research
BCG = Bacillus Calmette Guerin
CMF = Cytoxan, methotrexate, fluorouracil
CMFVP = Cytoxan, methotrexate, fluorouracil, vincristine, prednisone
ECF = Epirubicin, cisplatin, fluorouracil
FAC = Fluorouracil, adriamycin, Cytoxan
MER = Glycoprotein derived from exoskeleton of BCG

References

1. Beatson GT. On the treatment of inoperable cases of carcinoma of the mamma. Suggestion of a new method of treatment with illustrative cases. Lancet 1896; 2:104.
2. Fisher B. Lumpectomy compared with lumpectomy and radiation therapy for the treatment of intraductal breast cancer. N Engl J Med 1993; 328:1581–1586.
3. Fisher B. The science and politics of breast cancer: A personal perspective. Abstracts Chemotherapy Foundation Symposium, Innovative Cancer Therapy for Tomorrow. 1995; 34:40. Cancer Invest 1996; 14(1):40–42.
4. Fox MS. On the diagnosis and treatment of breast cancer. JAMA 1979; 241:489–494.
5. Carter CL, Allen C, Hensen DE. Relation of tumor size,

lymph node status, and survival in 24,740 breast cancer cases. Cancer 1989; 63:181–187.

6. Fisher ER. Prognostic and therapeutic significance of pathological features of breast cancer. Presented at NIH Consensus Development Conference on Adjuvant Chemotherapy and Endocrine Therapy of Breast Cancer, 1985. NCI Monogr 1986; 1:39–44.
7. McGuire WL, Clark GM. Prognostic factors and treatment decisions in axillary node-negative breast cancer. N Engl J Med 1992; 326:1756–1760.
8. Greenspan EM, Feiber M, Lesnick G, et al. Response of advanced breast carcinoma to the combination of the antimetabolite methotrexate and the alkylating agent thio-TEPA. J Mt Sinai Hosp 1963; 30:246–267.
9. Greenspan EM. Regression of metastatic hepatomegaly from mammary carcinoma. Cytotoxic combination chemotherapy with 5-fluorouracil. NY State J Med 1964; 64:2442–2449.
10. Greenspan EM. Combination cytotoxic chemotherapy in advanced disseminated breast carcinoma. J Mt Sinai Hosp 1966; 33:1–27.
11. Cooper RG. Combination chemotherapy in hormone resistant breast cancer. Proc Am Assoc Cancer Res 1969; 10:15.
12. Cooper RG, Holland JF, Glidewell O. Adjuvant chemotherapy of breast cancer. Cancer 1979; 44:793–798.
13. Fisher B, Redmond C, Fisher ER, et al. Systemic adjuvant therapy in treatment of primary operable breast cancer: National Surgical Adjuvant Breast and Bowel Project Experience. NIH Consensus Development Conference on Adjuvant Chemotherapy and Endocrine Therapy of Breast Cancer, 1985. NCI Monogr. 1986; 1:1–10.
14. Nissen-Mayer R, Host H, Kjellgren K, et al. Short perioperative versus long-term adjuvant chemotherapy: Perioperative chemotherapy: Rationale, risk and results. In: Metzger U, Largiader F, Senn JJ, eds. Recent results in cancer research. Springer-Verlag, Berlin 1985: 91–98.
15. Bonadonna G. Evolving concepts in the systemic adjuvant treatment of breast cancer. Cancer Res 1992; 52:2127–2137.
16. Greenspan EM. Inflammatory breast cancer. In: Ariel I, Cahan A, eds. Preneoplastic and early breast cancer. Williams and Wilkins, Baltimore, MD, 1993:280–292.
17. Early Breast Cancer Trialists Collaborative Group. Effects of adjuvant tamoxifen and cytotoxic therapy on mortality in early breast cancer. An overview of 61 randomized trials among 28,986 women. N Engl J Med 1988; 319:1681–1692.
18. Adler A, Stein JA, Ben-Efraim S. Immunocompetence, immunosuppression and human breast cancer. Cancer 1980; 45:2061–2083.
19. Krown SE, Pinsky CM, Wanebo HJ, et al. Immunologic reactivity and prognosis in breast cancer. Cancer 1980; 46:1746–1752.
20. Mandeville R, Lamoureux G, Legault-Poisson S, et al. Biological markers and breast cancer. Multiparametric study II. Depressed immunocompetence. Cancer 1982; 50:1280–1288.
21. Greenspan EM. Is there a relation between previous tuberculosis and long-term survival after solid tumor chemotherapy? Cancer Treat Rep 1982; 66:206–208.
22. O'Connell MJ, Ritts RE, Moertel CG. A controlled immunological evaluation of immunotherapy with BCG in patients with metastatic cancer. Cancer Clin Trials. 1981; 4:439–443.
23. Hortobagyi GI, Buzdar AU. The 10 year + results of combination chemotherapy for stage II and III breast cancer. Cancer Invest 1994; 12(1):A41–A42.
24. Weil M, Auclerc G, Borel Ch, et al. Neo-adjuvant chemotherapy in breast cancer study of 477 evaluable patients with primary breast cancers treated between 1980 and 1992. Ann Chir Plast Esthetique 1992; 37:663–669.
25. Jones AL, Smith IE, O'Brien D, et al. Phase II study of continuous infusion fluorouracil with epirubicin and cisplatin in patients with metastatic and locally advanced breast cancer: An active new regimen. J Clin Oncol 1994; 12:1259–1265.
26. Smith IE. Preoperative infusional chemotherapy for large early breast cancer: A very active new approach with high complete remission rates. Abstracts Chemotherapy Foundation Symposium, Innovative Cancer Therapy for Tomorrow. Cancer Invest 1996; 14(1):A42–A43.
27. Greenspan EM. Is BCG an "orphan drug" suffering from chemotherapists' overkill? Cancer Invest 1986; 4: 89–92.
28. Greenspan EM. Prior tuberculosis in long term survivors of breast and ovarian cancer from a New York oncology practice (1954–1974). Mt Sinai J Med 1985; 52:465–468.
29. Donegan WL, Kubo K, Shimuzu H, et al. Postsurgical survival of Japanese and American women with cancer of the breast: A cooperative study. Presented at the 13th International Cancer Congress, Seattle, Washington, Sept. 8–15, 1982.
30. Greenspan EM. Letter. Lancet. 1994; 344:8934.
31. Sudden fall in breast cancer death rates in England and Wales. Lancet. 1995; 345:1642.
32. Sainsbury R, Haward B, Rider L, et al. Round C: Influence of clinician workload and patterns of treatment on survival from breast cancer. Lancet. 1995; 345:1265–1270.
33. Skegg DCG. Breast cancer: Progress at last? Lancet 1995; 346:15.
34. Greenspan EM. Has the cure of breast cancer with chemotherapy become a reality. Cancer Invest 1994; 12: 98–100.
35. Greenspan EM. The cure of breast cancer by combination chemotherapy. Cancer Invest 1996;14(1):A70–73.
36. Anderson J, Nielsen J, Christensen L, et al. The significance of the frequency and occurrence of breast cancer for treatment: Essential aspects of new Danish studies of the natural history of breast cancer. Acta Chir Belg 1987; 87:103–107.
37. Stjernsward J. Decreased survival related to irradiation postoperatively in early operable breast cancer. Lancet 1974; 2:1284–1286.

Hypertension:

A Personal Odyssey

Milton Mendlowitz, M.D.

Abstract

This is a brief history of my personal experiences at the Mount Sinai Hospital and Medical School, preceding and interlaced with the research that eventually involved hypertension. My research on the hearts, kidneys, and lungs of animals was followed by work on clubbing of the fingers, hypertrophic osteoarthropathy, and myocardial infarction. This finally led to studying vascular resistance and reactivity in hypertension and other vascular diseases. I investigated and treated special causes of hypertension and developed proper testing and effective treatment of "essential" hypertension. We established the hypertension clinic and laboratory to accomplish these objectives. As for etiology, hereditary and environmental factors have become better defined, especially in "essential" hypertension. Study of the spontaneously hypertensive rat may clarify some hereditary factors in hypertension, although its relationship to human essential hypertension remains questionable. In addition, basic work on the physiology and molecular biology of uterine and vascular smooth muscle has shed some light on the complexities of restenosis after angioplasty, arteriosclerosis, left ventricular hypertrophy, and hypertension. The help and encouragement of teachers and colleagues throughout the years has made all this effort possible.

For those who know me now as a worker in molecular cardiology and remember me as a hypertension specialist, it might be of some interest to know how it all happened. In 1932 I came to Mount Sinai Hospital as what was then called a pathology intern and spent 6 months each with Dr. Gregory Shwartzman, of the Shwartzman phenomenon, and Dr. Paul Klemperer, the renowned pathologist. After my year of rotating internship, Dr. Sheppard Siegal and I, together with the assigned nurses, took charge of 80 inpatients in two wards, each of us for 2 years. We were called house physicians rather than residents. With two helping interns, we wrote up the history, physical examination, progress and discharge notes on all patients and we also left orders. We even performed urinalyses and blood counts. It was a disgrace to have an attending physician suggest a test or treatment that had not already been ordered or carried out. Antibiotics had not yet been discovered. (The first antibiotic, prontosyl, a sulfa drug, was brought from Europe by one of the attending physicians.) Either Dr. Siegal, myself, or both of us made rounds three times a day: morning, afternoon (with attending physicians), and at midnight. We were given our meals and quarters but were not paid. No one could be appointed if they were married.

During this period I came to know many of the older physicians. Among them were Dr. Emanual Libman, who described subacute bacterial endocarditis (1); Dr. George Baehr, who worked on the kidney in lupus erythematosis (2); Drs. Crohn, Ginzberg, and Oppenheimer (3) who described what was later known as Crohn's disease; and many others. Clinical research was a Mount Sinai tradition and many physicians combined this with their clinical practices. Some of these investigations were in bacteriology, saline infusion, and blood transfusion, but a great deal were in pathology.

This work was aided by a grant from The Louis and Harold Price Foundation, Inc.

Address reprint requests to Milton Mendlowitz, M.D., 1200 Fifth Avenue, Apt. 15A, New York, NY 10029.

Dr. Louis Gross was in charge of cardiac pathology at that time and he accepted me in his laboratory for work on pathologic physiology. My first papers with Drs. Gross and Schauer were on the decreased cardiac output (cardiogenic shock) after coronary ligation in dogs (4). We used a special knot, the double carrick bend, to occlude the left anterior descending artery in the anesthetized dog without having to open the chest at the time of ligation (5).

After that, in Chicago, I studied heart failure in the isolated dog heart (6) and the effect of digitalis on the preparation (7) with the physiologist, Dr. Louis Katz. Dr. Katz later stated that this work sustained him at meetings for 10 years thereafter. In his laboratory, I also investigated experimental renovascular hypertension (8) and experimental subacute bacterial endocarditis (9), in all dogs, with Dr. Meyer Friedman [who later wrote on coronary type A and type B personality (10)]. I also introduced massive pulmonary emboli in anesthetized dogs, causing another type of shock with decreased cardiac output (11). Studies on the isolated heart preparation were also reported (12).

I was then advised to do clinical research for a year, to study the opposite of a low cardiac output: either hyperthyroidism or possibly clubbed fingers and hypertrophic osteoarthropathy. Sir Thomas Lewis in London had established clinical research as a new discipline and he accepted me for the clubbing project. I traveled all over London testing patients with chest disease and people with hereditary clubbing; and I finally demonstrated that in acquired developing clubbing, finger tip blood flow was increased (13). Also, we eventually invented the digital thermal flowmeter (13, 14). In New York, at Mount Sinai, Dr. Alan Leslie and I produced hypertrophic osteoarthropathy by anastomosing a dog's left auricle to the pulmonary artery (15). I wrote a complete review of the subject for *Medicine* (16) which, I subsequently learned, was used for teaching in medical schools all over the country. All this work was done with the support of fellowships.

World War II broke out and it was not long before I was in the Army. I wrote and published a few of my papers while I was in the Army, such as one on left pectoralis minor strain as a cause of left anterior chest pain in soldiers (17) and, in the Phillipines, one on blood volume measured by dilution after administration of plasma in shock due to blood loss (18).

After the war, on returning to New York and Mount Sinai, my minuscule practice left time for research and I became interested in coronary disease and hypertension. I established the first Hypertension Clinic at the Hospital. With the help of fellowships and finally Heart Association and government grants, I began to use the digit as a paradigm for the systemic circulation because I could measure both pressure and flow while perfusing various agonists.

At that time I obtained permission to study patients who had myocardial infarction with the digital calorimeter, using an arm warm-water bath to release sympathetic nerve tone. Because there was no space for a research laboratory on the floor, all the apparatus had to be stored in the nurses' clean-up area. As in the dog experiments, blood flow was decreased (19).

Dr. Alexander Gutman was then appointed the first full-time Chief of Medicine at Mount Sinai. He had been associated previously with Presbyterian Hospital of Columbia University. At that time Mount Sinai was affiliated with Columbia University, and Dr. Gutman had worked at the Columbia Division of the Goldwater Memorial Hospital situated on what was then called Welfare Island. Dr. David Siegal replaced Dr. Gutman at Goldwater and Dr. Gutman urged me to attend seminars there. Dr. Robert Loeb, the Chief of Medicine at Presbyterian Hospital, participated in those meetings. It was because of this association, that N. Naftchi and I were able to study a Goldwater patient who had essential hypertension. To our astonishment we found that after release of sympathetic tone by body heating, the patient did not demonstrate an increase in such tone, which we expected, but instead had a huge increase in vascular reactivity to the infused sympathetic neuroeffector, norepinephrine. Some but not all of this was caused by vascular hypertrophy and hyperplasia, as we learned from subsequent studies (20, 21).

Soon it became clear that hypertension could have many different causes. Because we were investigating the catecholamines, we became very interested in various types of pheochromocytomas (22), most of which could be cured surgically by adrenalectomy (23). Dr. Stanley Gitlow in our laboratory became an authority on the subject. In Cleveland, Page and Helmer (24) and, in Buenos Aires, Braun-Menendez et al. (25) discovered renin and Haber and Carlson (26) clarified the renin cascade ending in angiotensin II. It was known that parenchymal nephritis could be associated with hypertension, but attention was directed toward renovascular hypertension because of the possibility of surgical cure (27). Again, some patients were operated on without effect (28) until the indications for invasive intervention were clarified (29). Surgery was eventually partially replaced by angioplasty (30).

The adrenal cortical causes of hypertension were also studied. Before this, Dr. Harvey Cushing (31) had described the clinical features of hyperglucocorticoid hypertension. Some of these cases originated when a pituitary adenoma overproduced adrenocorticotropic hormone (ACTH), causing adrenal cortical hyperplasia. Other cases of Cushing syndrome were caused by adenomas or carcinomas of the adrenal cortex (32). Depending on the testing, cure was either pituitary ablation or adrenalectomy. My University of Michigan Medical School classmate, Dr. Jerome Conn then described the syndrome that bears his name (33). This syndrome was caused by adrenal adenoma or hyperplasia secreting aldosterone which produced sodium retention and also hypertension. This too was usually curable by adrenalectomy.

Acute hypertension could also occur with brain tumor (34) or cerebral hemorrhage (35). Anxiety was often associated with transient hypertension and this could be caused by the simple presence of a physician ("white-coat" hypertension) (36). Pseudohypertension was described in the elderly and obese due to error in the indirect measurement of blood pressure (37) because of a fat arm or "hardening" of the arteries caused by arteriosclerosis (38).

All these special cases of hypertension represented only about 10% or less of the bulk of hypertensive patients. What is more, hypertension afflicted nearly 30% of the entire population (39). This meant that we were dealing with an enormous problem, most of the etiology of which was obscure. After excluding specific causes, the remaining patients were defined as those with "essential" or "primary" hypertension, the causes of which were unknown (40).

In the 1940s the subject of hypertension was in an unusually confused state. Because various proposed remedies, such as phenobarbital, had been ineffective, authorities at two of the major academic institutions in New York refused to treat hypertensive patients with any drugs. Some patients were treated experimentally with saline infusions by one of these authorities. Sympathectomy was in vogue in Boston (41) and in Michigan (42), but many patients were unresponsive. Even when reserpine was shown to be an effective antihypertensive drug, it took a long time for it to be accepted in New York, whereas in Boston and Cleveland drug therapy had already begun (43, 44). In our own Hypertension Clinic, we became committed to drug therapy (45) to prevent stroke and especially malignant hypertension with renal failure, which was not an uncommon effect of hypertensive disease at that time (46).

Historically effective treatment of primary hypertension began with reserpine (44). After reserpine, the diuretics became popular (47, 48), especially hydrochlorothiazide. Then the beta blockers (49) were in vogue followed by the alpha-norepinephrine inhibitors (50) and blockers (51). More recently, because the angiotensin system involvement in hypertension could be tested (52), the angiotensin I converting enzyme (ACE) inhibitors (53) became useful and finally, the calcium channel blockers (54) and the angiotensin antagonists (55) were made available. All these drugs reduced blood pressure, but side effects were not uncommon. Reserpine could depress the brain (56), the diuretics could produce hypokalemia (57), the beta blockers might depress cardiac output and produce cold hands (58), the ACE inhibitors could produce cough (59), the calcium channel blockers could affect the heart (60), and the list goes on. Overtreatment could reduce blood pressure too much and could make the patient susceptible to a myocardial infarction (61). Even conservative treatment could produce electrocardiographic changes such as T-wave inversion without any associated symptoms (62). The drugs did, however, prevent stroke, malignant nephrosclerosis, and renal failure to some extent, thus reducing morbidity and mortality in patients with essential hypertension (63).

After our initial observation at Goldwater our laboratory at Mount Sinai continued to study vascular reactivity. Aldosterone with salt increased it (64). Prednisone (65) had no effect on reactivity in patients with essential hypertension but increased such reactivity in normotensive individuals. The role of salt is still somewhat obscure. Saline infusions paradoxically increased urinary sodium excretion in the patient with essential hypertension (66). Some patients with hypertension were salt sensitive and some were not (67). Diuretics increased sodium and potassium urinary excretion and decreased blood pressure and reactivity (68). A ouabain-like substance that could affect sodium metabolism was found in the blood of patients with hypertension (69). In addition, atrial natriuretic hormone was found to be involved in some hypertensive disease (70). Also, endothelin (71), a vasoconstrictor, and endothelial relaxing factor, a vasodilator, were discovered (72). The renin-angiotensin system was found to be autocrine in vascular smooth muscle (73) as well as endocrine (74). This led to more extensive use of ACE inhibitors and angiotensin II antagonists in the treatment of primary hypertension.

Vascular reactivity was also found to be increased in the offspring of parents with essential hypertension before the children developed hypertension (75). Also, hypertension could develop in each of monozygotic twins even if they were reared apart from each other (76). So, there were certainly hereditary factors in essential hypertension on which environmental factors were superimposed.

The environmental factors became clearer when the 24-hour blood pressure apparatus became available (77). This disclosed anxiety hypertension and differentiated it from true essential hypertension. The patients with "white coat" syndrome did not seem to develop progressive hypertensive disease (78). Another aid in this respect was the echocardiogram and thallium imaging which could reveal early left ventricular hypertrophy and some of its associated dynamic effects (79). This could place the patient in the progressive hypertensive group and warrant antihypertensive treatment.

Patients with parenchymal nephritis also had hypertension and, in these cases, it was often difficult to separate cause from effect (80). Did essential hypertension make an individual more susceptible to parenchymal nephritis? These possibilities still need to be sorted out.

Collateral effects of our work on hypertension were measurements of vascular resistance in the big toe, where vascular resistance was increased in peripheral vascular disease (81) and also in diabetes mellitus (82). It was increased in the finger in advanced Raynaud's disease (83) and polycythemia (84). Vascular resistance and reactivity were also studied in the hypertension of pregnancy (85).

Besides research, rounds, and private practice, there was always a succession of meetings. There were the annual hypertension meetings in Cleveland at which I became chairman for a year. Also there were Heart Association meetings, international hypertension meetings in addition to the Clinical Investigation and American Medical Association meetings. Locally, at Mount Sinai, I was a member of the Association of Attending Physicians and eventually became its president. As president of the Association, at the Medical Board meetings led by Dr. George James, along with others, I was involved in establishing the Mount Sinai School of Medicine, by influencing the Attending Physicians to support the project. In addition, my patients, Joe Lowe and Louis Price established one of the first Chairs in Medicine, which I occupied until becoming emeritus.

Meanwhile, Dr. Lawrence Krakoff was recruited for the Hypertension Division of the Department of Medicine, including the clinic and the laboratory. He had been working at Presbyterian Hospital with Dr. John Laragh and, in our laboratory, he established testing for the renin angiotensin system in hypertension. This helped clarify which patients with renovascular hypertension could be helped or even cured by intervention and which could not. It also helped to identify corticosteroid hypertension (86) and to determine which drug therapy to use in difficult cases of primary hypertension. With my age-related retirement, Dr. Krakoff took over the hypertension clinic and laboratory until he retired from Mount Sinai to become Chief of Medicine at Englewood Hospital in New Jersey. Since then Dr. Robert Phillips, a cardiologist interested in the cardiac effects of high blood pressure (87) has taken the helm.

Upon retirement from medical practice I became a laboratory volunteer in the department of the physiologist, Dr. Irving Schwartz. Dr. Diana Gazis and I studied smooth muscle. We measured uterine contraction and blood pressure in female rats in response to almost every agonist known to affect such smooth muscle responses (88–90).

As smooth muscle biochemistry became better understood (91) it seemed that only the G_s protein and phospholipase C could increase contraction as well as growth in vascular smooth muscle, and these two functions were increased in essential hypertension. If there was overregulation of either of these proteins it might throw light on one possible hereditary factor in essential hypertension (**figure**).

I chose to study phospholipase C with the help of Dr. Mark Taubman, a gene expert, who had been recruited for the Mount Sinai School of Medicine. Although there were six isoforms of the gene, the major one we isolated from rat vascular smooth muscle was the alpha-isoform. This gene for phospholipase C was constitutively expressed (92). In other words, it was abundantly and continuously available and apparently not regulated. Undoubtedly the geneticists of the future will unravel the biochemical changes in vascular smooth muscle, including DNA, RNA, and protein, which constitute the hereditary factors in essential hypertension. I have also grown and freeze-stored aortic smooth muscle cells from Wistar Kyoto (WKY) and spontaneously hypertensive rats (SHR). Like the Japanese (93), I have seen the SHR cells grow faster than those from the WKY rats. The Japanese (94) have also demonstrated increased expression of the c-*myc* protooncogene in the SHR heart and vascular smooth muscle

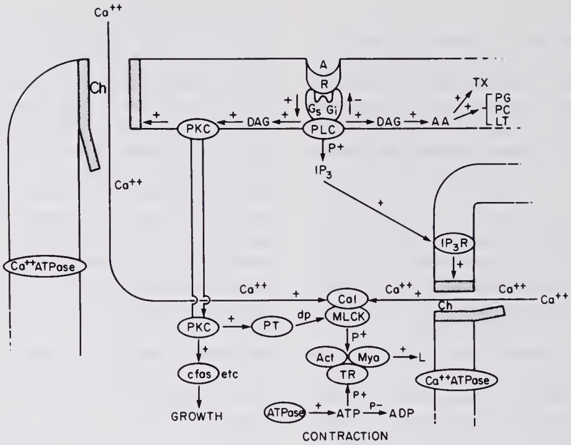


Fig. The phospholipase C cascade. A, agonist; R, receptor; G_s, stimulatory; G_i, inhibitory; PLC, phospholipase C; DAG, diacylglycerol; PKC, protein kinase C; AA, arachidonic acid; Tx, thromboxan; PG, prostaglandin; PC, prostacyclin; LT, leucotriene; Ca^{++}, calcium ion; Ca^{++} ATPase, ion pump; Ch, channel; IP_3, phosphoinositide; IP_3R, phosphoinositide receptor; Cal, calmodulin; MLCK, myosin light chain kinase; cfos etc., growth proteins; Act, actin; Myo, myosin; TR, tropomyosin; L, latch state; PT, phosphatase; ATP, adenosine triphosphate; ADP, adenosine diphosphate; P−, loss of phosphorus; P+, phosphorylation; dp, dephosphorylation; +, positive chemical change. The larger membrane is the cellular membrane and the smaller one is the sarcoplasmic reticular membrane. Some intermediate and collateral steps are omitted. Vascular dilatation (not shown) and growth inhibition (not shown) are also chemically active and positive. They oppose the PLC cascade.

cells than in those from the WKY rats. The relation of these differences to human essential hypertension is questionable, of course, but the differences deserve further study.

Meanwhile, with others in Dr. Taubman's laboratory, I am studying balloon angioplasty in the aorta or carotid artery of the rat, where various treatments may modify the resulting arteriosclerosis. This implicates the clotting cascade, initiated by "tissue factor" (95) and other genes and proteins. The ultimate goal of this work is the reduction of restenosis after coronary angioplasty and modification of the process of human arteriosclerosis. The multiple factors in the development of such sclerosis include hypertension.

References

1. Libman E, Sacks B. A hitherto undescribed form of valvular and mural endocarditis. Arch Intern Med 1924; 33:701–737.
2. Baehr G, Klemperer P, Schifrin A. DIffuse disease of peripheral circulation usually associated with lupus erythematosis and endocarditis. Trans Assoc Am Physicians 1935; 50:139–155.
3. Crohn BB, Ginzburg L, Oppenheimer GD. Regional Ileitis. A pathological and clinical entity. JAMA 1932; 99: 1323–1329.
4. Gross L, Mendlowitz M, Schauer G. Hemodynamic studies in experimental coronary occlusion. Open chest experiments. Am Heart J 1937; 13:647–663.
5. Gross L, Mendlowitz M, Schauer G. Hemodynamic studies in experimental coronary occlusion. Closed chest experiments. Am Heart J 1937; 13:664–676.
6. Katz LN, Mendlowitz M. Heart failure analyzed in the isolated heart circuit. Am J Physiol 1938; 122:262–273.
7. Katz LN, Mendlowitz M, Kaplan HA. The action of digitalis on the isolated heart. Am Heart J 1938; 16:149–158.
8. Katz LN, Mendlowitz M, Friedman M. Study of the factors concerned in renal hypertension. Proc Soc Exp Biol Med 1938; 37:722–724.
9. Friedman M, Katz LN, Howell K, Mendlowitz M, Lindner E. Experimental endocarditis due to streptococcus viridans. Arch Intern Med 1938; 61:95–118.
10. Friedman M, Ghandour G. Medical diagnosis of type A behavior. Am Heart J 1993; 126:607–618.
11. Mendlowitz M. Experimental pulmonary embolism. J Thor Surg 1938; 8:204–218.
12. Katz LN, Mendlowitz M. The isolated heart circuit. Proceedings of the International Physiology Congress, Zurich, 1938.
13. Mendlowitz M. Some observations on clubbed fingers. Clin Sci 1938; 3:387–400.
14. Mendlowitz M, Feitelberg S. A recording thermal digital flowmeter. J Appl Physiol 1956; 8:671–673.

15. Mendlowitz M, Leslie A. The experimental simulation in the dog of the cyanosis and hypertrophic osteoarthropathy which are associated with congenital heart disease. Am Heart J 1942; 24:141–152.
16. Mendlowitz M. Clubbing and hypertrophic osteoarthropathy. Medicine 1942; 21:269–306.
17. Mendlowitz M. Strain of the pectoralis minor muscle, an important cause of precordial pain in soldiers. Am Heart J 1945; 30:123–125.
18. Mendlowitz M. Blood volume determination. Bull US Army Med Dep 1948; 8:58–61.
19. Mendlowitz M. The digital blood flow, arterial pressure and vascular resistance in arterial hypertension and in coronary thrombosis. J Clin Invest 1942; 21:539–546.
20. Mendlowitz M, Naftchi N. The work of digital vasoconstriction produced by infused norepinephrine in primary hypertension. J Appl Physiol 1958; 13:247–251.
21. Mendlowitz M, Torosdag A, Sharney L. The force and work of digital anteriolar vasoconstriction in hypertension. J Appl Physiol 1957; 10:436–446.
22. Gitlow SE, Khassis S, Cohen G, Mendlowitz M, Sha J. Diagnosis of pheochromocytoma by determination of urinary 3-methoxy, 4-hydroxy mandelic acid. J Clin Invest 1960; 39:221–226.
23. Gitlow SE, Mendlowitz M, Wolf RL. The diagnosis of pheochromocytoma. J Mt Sinai Hosp 1961; 28:159–179.
24. Page IH, Helmer DM. A crystalline pressor substance (angiotonin) resulting from the reaction between renin and renin activator. J Exp Med 1940; 71:29–42.
25. Braun-Menendez E, Fasciolo JC, Leloir LF, Munoz JM. La substancia hypertensora de la sangre del riñon isquimiado. Rev Soc Argent Biol 1939; 15:420–425.
26. Haber E, Carlson W. The biochemistry of the renin-angiotensin system. In: Genest J, Kuchel O, Hamet P, Cantin M, eds. Hypertension. New York: McGraw Hill. 1977:171–184.
27. Novick AC. Renovascular hypertension surgical treatment. In: Laragh JG, Brenner BM, eds. Hypertension. New York: Raven Press. 1990; 2:1561–1571.
28. Genest J, Cartier P, Roy P, et al. Renovascular hypertension. In: Genest J, Kuchel O, Hamet P, Cantin M, eds. Hypertension. 2nd ed. New York: McGraw Hill. 1983: 1007–1034.
29. Pickering TG. Renovascular hypertension. In: Laragh JH, Brenner BM, eds. Hypertension. New York: Raven Press 1990; 2:1539–1571.
30. Grüntzig A, Kuhlmann U, Vetter W. Treatment of renovascular hypertension with percutaneous transluminal dilatation of a renal artery stenosis. Lancet 1978; 1: 801–802.
31. Cushing H. The basophil adenomas of the pituitary body and their clinical manifestations (pituitary basophilism). Bull Johns Hopkins Hosp 1932; 50:137–195.
32. Biglieri EG, Irony I, Kater CE. Adrenocortical forms of human hypertension. In: Laragh JH, Brenner BM, eds. Hypertension. New York: Raven Press. 1990; 2:1609–1623.
33. Conn JW. Primary aldosteronism, a new clinical syndrome. J Lab Clin Med 1955; 45:3–17.
34. Bell GM, Doig A. Hypertension and intracranial tumor. In: Robertson JS, ed. Handbook of hypertension. Vol. 2: Clinical aspects of secondary hypertension. Amsterdam: Elsevier. 1983:291–303.
35. Britton M, Carlsson A, deFaire U. Blood pressure course in patients with acute stroke and matched controls. Stroke 1986; 17:861–864.
36. Pickering TG, James GD, Boddie C, Harshfield GA, Laragh JH. How common is the white coat syndrome? JAMA 1987; 259:225–228.
37. Trout KW, Bertrand CA, Williams MH. Measurement of blood pressure in obese persons. JAMA 1956; 162:970–971.
38. Sprague DL, Kim DI. Pseudohypertension due to Mönkeberg's arteriosclerosis. Anesth Analg 1978; 57:588–589.
39. Kaplan NM. Hypertension in the population at large. In: Clinical hypertension. 6th ed. Baltimore, MD: Williams and Wilkins Co. 1994:1–22.
40. Kirkendall WM. Essential hypertension. In: Genest J, Kuchel O, Hamet P, Cantin M, eds. Hypertension. New York: McGraw Hill. 1983:732–753.
41. Smithwick RH. Surgical treatment of hypertension. Am J Med 1948; 4:744–759.
42. Peet NM. Results of subdiaphragmatic splanchnicectomy for arterial hypertension. Am J Med Sci 1947; 236:270–276.
43. Wilkins RW, Judson WE. Use of rauwolfia serpentina in hypertensive patients. N Engl J Med 1953; 248:48–53.
44. Dustan HP, Schneckloth RE, Corcoran AC, Page IH. The effectiveness of long-term treatment of malignant hypertension. Circulation 1958; 18:644–651.
45. Mendlowitz M. The treatment of hypertension with drugs. Ann Intern Med 1953; 39:999–1013.
46. Kincaid-Smith P. Malignant hypertension: Mechanisms and management. Pharmacol Ther 1981; 9:245–269.
47. Veterans Administration Cooperative Study Group on Antihypertensive Agents. Low doses vs. standard dose of reserpine, a randomizing double-blind multiclinic trial in patients taking chlorothalidone. JAMA 1982; 248:2471–2477.
48. Freis ED, Wilson IM, Parrish AE. Enhancement of antihypertensive activity with chlorothiazide. Presented at the 19th annual meeting of the American Heart Association, Chicago, Illinois, October 28, 1957.
49. Black JW, Stephenson JS. Pharmacology of a new adrenergic beta-receptor blocking compound (Nethalide). Lancet 1962; 2:311–314.
50. Van Zwieten PA. The central action of antihypertensive drugs, mediated via central alpha-receptors. J Pharmacol 1973; 25:89–95.
51. Van Zwieten PA. Alpha-adrenoceptor-blocking agents in the treatment of hypertension. In: Laragh JH, Brenner BM, eds. Hypertension. New York: Raven Press. 1990: 2233–2249.
52. Sealy JE, Laragh JH. The renin-angiotensin-aldosterone system for normal regulation of blood pressure and sodium and potassium homeostasis. In: Laragh JH, Brenner BM, eds. Hypertension. New York: Raven Press. 1990:1287–1317.
53. Gavras H, Brenner HR, Turini GA, et al. Antihypertensive effect of oral angiotensin converting enzyme inhibition. SQ 14225 in man. N Engl J Med 1978; 298:991–995.
54. Aoki K, Kondo S, Moshizuki A. Antihypertensive effects of cardiovascular Ca^{2+}-antagonists in hypertensive patients in the absence and presence of beta-adrenergic blockade. Am Heart J 1978; 96:218–227.
55. Soffer BA, Wright JT Jr, Pratt JH, Wiens B, Goldberg AI, Sweet CS. Effects of Losartan on a background of hydrochlorothiazide in patients with hypertension. Hypertension 1995; 26:112–117.
56. Prisant LM, Spruill WJ, Fincham JE, Wade WE, Carr AA, Adams MA. Depression associated with antihypertensive drugs. J Fam Pract 1991; 33:481–485.

57. Morgan DB, Davidson C. Hypokalemia and diuretics: An analysis of publications. Br Med J 1980; 280:1905–1908.
58. Zacharias FJ, Cowen KJ, Vickens J, Wall BC. Propranolol in hypertension. A study of long-term therapy 1964–1970. Am Heart J 1972; 83:755–761.
59. Morice AH, Lowry R, Brown MJ, Higgenbottam T. Angiotensin converting enzyme and the cough reflex. Lancet 1987; 2:1116–1118.
60. Walker PC, Inman WHW. Diltiazem and heart block. Lancet 1989; 1:617.
61. Stewart IMcDG. Relation of reduction in pressure to first myocardial infarction in patients receiving treatment for severe hypertension. Lancet 1979; 1:861–865.
62. Mendlowitz M. Treatment of uncomplicated essential hypertension with emphasis on the associated electrocardiographic (ECG) changes. Mt Sinai J Med 1983; 50: 388–392.
63. Freis ED. Effects of treatment on morbidity in hypertension. Results in patients with diastolic blood pressure averaged 115–129 mm Hg. Veterans Administration Cooperative Study Group on Antihypertensive Agents. JAMA 1967; 202:1028–1034.
64. Mendlowitz M, Naftchi NE, Bobrow EB, Wolf RL, Gitlow SE. The effect of aldosterone on electrolytes and on digital vascular reactivity to 1-norepinephrine in normotensive, hypertensive and hypotensive subjects. Am Heart J 1963; 65:93–101.
65. Mendlowitz M, Naftchi N, Weinreb HL, Gitlow SE. The effect of prednisone and prednisolone on reactivity of the digital blood vessels to 1-norepinephrine in normotensive and hypertensive subjects. J Appl Physiol 1961; 16:89–94.
66. Green DM, Wedell MD, Wald MH, Learned B. The relation of water and sodium excretion. Circulation 1952; 6:919–924.
67. Dahl LK, Heine M, Tassimari L. Effects of chronic salt ingestion. Evidence that genetic factors play an important role in susceptibility to experimental hypertension. J Exp Med 1962; 115:1173–1190.
68. Mendlowitz M, Naftchi N, Gitlow SE, Wolf RL. The effect of chlorothiazide and its congeners on the digital circulation in normotensive subjects and in patients with essential hypertension. Ann NY Acad Sci 1960; 88:964–974.
69. Vardaman MB Jr, Haddy FJ. Circulating natriuretic factors in hypertension. In: Laragh JH, Brenner BM, eds. Hypertension. New York: Raven Press. 1990:939–954.
70. deBold AJ, Borenstein HB, Veress AT, Sonnenberg H. A rapid and potent natriuretic response to intravenous injection of atrial myocardial extract in rats. Life Sci 1981; 28:89–94.
71. Yanagisawa M, Kurchara H, Kimura S, Tomobe Y, Kobayashi M, Mitsui Y, Yasake Y, Goto K, Masaki T. A novel potent vasoconstrictor peptide produced by vascular endothelial cells. Nature 1988; 132:411–415.
72. Vanhoutte PM, Rubanyi GM, Miller VM, Houston DS. Modulation of vascular smooth muscle contraction by the endothelium. Annu Rev Physiol 1986; 48:307–320.
73. Dzau VJ. Significance of the vascular renin-angiotensin pathway. Hypertension 1986; 8:553–559.
74. Goldblatt H, Katz YJ, Lewis HA, Richardson E. Studies in experimental hypertension. Production of persistent elevation of systolic blood pressure by means of renal ischemia. J Exp Med 1934; 59:347–379.
75. Hofman A, Grobbe DE, Schalekamp MADH, eds. The early pathogenesis of primary hypertension. Amsterdam: Escerpta Medica, 1987.
76. Hong Y, deFaire U, Heller DA, McClearn GE, Pedersen N. Genetic and environmental influences on blood pressure in elderly twins. Hypertension 1994; 24:663–670.
77. Littler WA, Watson RDS. Circadian variation in blood pressure. Lancet 1978; 1:995–996.
78. Cavallini MC, Roman MJ, Pickering TG, Schwartz JE, Pini R, Devereux RB. Is white coat hypertension associated with arterial disease or left ventricular hypertrophy? Hypertension 1995; 26:413–419.
79. Phillips RA. The cardiologists' approach to evaluation and management of the patient with essential hypertension. Am Heart J 1993; 126:648–666.
80. Krakoff LR, Phillips RA, Elson HB. New directions for hypertension therapy. Compr Ther 1987; 13:61–68.
81. Mendlowitz M, Abel H. Great toe calorimetry in peripheral vascular disease. Circulation 1951; 4:120–122.
82. Mendlowitz M, Grossman EB, Alpert S. Decreased hallucal circulation as an early manifestation of vascular disease in diabetes mellitus. Am J Med 1953; 15:316–327.
83. Mendlowitz M, Naftchi N. The digital circulation in Raynaud's disease. Am J Cardiol 1959; 4:580–584.
84. Mendlowitz M. The effect of anemia and polycythemia on digital intravascular blood viscosity. J Clin Invest 1948; 27:565–571.
85. Mendlowitz M, Altchek A, Naftchi N, Spark R. Digital vascular reactivity to 1-norepinephrine in the second trimester of pregnancy as a test for latent essential hypertension and "toxemia." Am J Obstet Gynecol 1961; 81:643–652.
86. Krakoff LR. Glucocorticoid excess syndromes causing hypertension. Cardiol Clin 1988; 6:537–545.
87. Phillips RA. Etiology, pathophysiology and treatment of left ventricular hypertrophy: Focus on severe hypertension. J Cardiovasc Pharmacol 1993; 21(suppl 2):S55–S62.
88. Gazis D, Gonzales G, Mendlowitz M. Verapamil blockade of pressor and uterine responses to the AVP/OT analog, oxypressin. Am J Physiol 1988; 254:R75–R77.
89. Gazis D, Gonzales G, Mendlowitz M. Adrenergic interactions in uterus and vascular smooth muscle in rats in vivo. Can J Physiol Pharmacol 1989; 67:1586–1590.
90. Gazis D, Gonzales G, Mendlowitz M. In vivo simultaneous comparison of pressor and uterine responses to a single agonist, (oxypressin) in the estrus rat. Can J Physiol Pharmacol 1987; 65:6–11.
91. Berridge MJ. Inositol triphosphate and diacyl glycerol: two interesting second messengers. Annu Rev Biochem 1987; 56:159–193.
92. Griendling KK, Taubman MB, Akers M, Mendlowitz M, Alexander RW. Characterization of phosphatidyl-inositol-specific phospholipase C from cultured vascular smooth muscle cells. J Biochem 1991; 266(23):15498–15504.
93. Yamori Y, Igawa T, Tagami M, et al. Humoral trophic influences on cardiovascular structural changes in hypertension. Hypertension 1984; 6(suppl III):27–32.
94. Negoto N, Inariba H, Inoue T, Kanayama Y, Takeda T. Expression of c-myc proto-oncogene in hearts and cultured smooth muscle cells of spontaneously hypertensive rats. J Hypertens 1988; 6(suppl 4):S128–S130.
95. Marmur JD, Rossikhina M, Guha A, et al. Tissue factor is rapidly induced in arterial smooth muscle after balloon injury. J Clin Invest 1993; 91:2253–2259.

History of the Department of Dermatology of The Mount Sinai Hospital

Douglas D. Altchek, M.D.

Abstract

The Department of Dermatology at Mount Sinai has a long, distinguished history. Founded in 1891, the department started in an era of keen diagnosticians and clinicians. By the middle of the twentieth century, it had evolved as both a clinical and investigational department at the forefront of both dermatology and syphilology. Today, it excels in both clinical and research areas and is on the cutting edge in dermatologic surgery.

Dermatology and Mount Sinai have had a long relationship dating to biblical times. In fact, dermatology began at Mount Sinai. In the book of Leviticus, Chapter XIII, Verse 3, numerous references are made to "leprosy" or "tzaárat." (1) (Fig. 1) (All illustrations appear at the end of the article.)

> 1. And the Lord spoke unto Moses and unto Aaron, saying: 2. When a man shall have in the skin of his flesh the plague of leprosy, then he shall be brought unto Aaron the priest, or unto one of the sons of the priest. 3. And the priest shall look on the plague in the skin of the flesh; and if the hair in the plague be turned white, and the appearance of the plague be deeper than the skin of the flesh, it is the plague of leprosy; and the priest shall look on him, and pronounce him unclean.

Although leprosy did exist in biblical times, it is unlikely that many of the references to leprosy were really leprosy. It is probable that many chronic and inflammatory skin diseases such as acne vulgaris, psoriasis, and eczema were confused with leprosy. During the Middle Ages, it is likely that many "leper houses" contained numerous patients with psoriasis and eczema who may very well have acquired leprosy as a result of being confined to these houses in prolonged contact with true lepers.

At the time Jew's Hospital (later renamed Mount Sinai Hospital) was founded in 1852, there was no formal dermatology department (Fig. 2). Indeed, medical education in general in the United States was quite chaotic (2). In 1852 alone, 53 medical schools were established in the United States. Numerous incompetent "doctors" were practicing, often on haphazard theory. A medical license could be obtained in less than 6 months after attending a few lectures. In 1852, sepsis and antisepsis were practically unknown; doctors were often the carriers of disease.

New York City had an epidemic of typhus in 1842 and of Asiatic cholera in 1849. Violence, criminality, and immorality were widespread in the city. Prisoners of both sexes and of all ages were often herded into common pens, further spreading disease and pestilence (2). Syphilis affected nearly 10% of the population.

Although dermatology as a specialty was not officially recognized by the American Board of Specialties until 1938, when the American Board of Dermatology and Syphilology was established, the field had long been recognized in Europe. The fields of dermatology and syphilology were united

---

From the Department of Dermatology, Mount Sinai Medical Center, New York. Address reprint requests to Douglas Altchek, M.D., 115 East 61st Street, New York, NY 10021.

The reproduction of photographic illustrations was supported by a grant from Dermik Laboratories.

because of the numerous dermatologic manifestations of syphilis.

By the time Mount Sinai Hospital moved to its second location at Lexington Avenue between 67th and 68th Streets in 1882, there was a clear need for a department of dermatology (Fig. 3). At the time, many dermatologic conditions were treated with preparations that were derivatives of arsenic, mercury, sulfur, and tar (3). Psoriasis was treated with carbolic acid in glycerine, tar, oil of cade, or mercury ointment. Eczema was treated with 10% oil of cade (Fig. 4). Basal cell carcinoma was treated by arsenic paste or cautery (Fig. 5). For syphilis ("the great imitator"), as the saying went, "One night with Venus, and the rest of your life with mercury."

In 1891, Sigmund Lustgarten was appointed consulting dermatologist to the hospital (Fig. 6). He was a student of the great European dermatologists, von Hebra and Kaposi. He trained not only in dermatology, but also in general medicine, chemistry, and pathology. He came to the United States from his native Vienna in 1889 and became active in medical circles. He quickly gained a reputation as a skilled clinician and diagnostician. He did not limit his diagnostic work to dermatology alone, however, and he was frequently called to attend the general wards. At the time, diagnosis depended almost entirely on clinical knowledge.

On one famous occasion, he made the diagnosis of leprosy in a patient on the general wards. His diagnoses were usually preceded by a mysterious approach. In this case, he asked to have a "pimple" pricked on the lip of the patient. He made a smear for tubercle bacilli, although everyone was sure this was not a case of tuberculosis. The smear proved the patient had leprosy (4).

On another occasion, he visited a patient who had broken out with a fiery, red rash, which was diagnosed as scarlet fever. The patient was put in isolation. Dr. Lustgarten examined the patient, who was a syphilitic under treatment. He diagnosed the rash as a case of overdose of mercury, which had been given to treat syphilis.

Dr. Lustgarten was a man of extreme discretion and tact. In going through a hotel lobby one day with another dermatologist, Dr. Richard Hoffman, Hoffman saw and greeted a man they both knew. Dr. Lustgarten did not acknowledge the man. When Hoffman pointed out that he was someone they both knew, Lustgarten explained that he rarely greeted people in public, because his reputation as a syphilis expert might embarrass some of his acquaintances. He was the first to give salvarsan as a treatment for syphilis.

Like many of his colleagues, Dr. Lustgarten was a man of culture and erudition, a lover of music, and a connoisseur of art. He was also a good draftsman and a collector of fine engravings and etchings. Soft spoken and dignified, he was considered a man of great intellectual stature and commanded universal respect.

On Lustgarten's death in 1910, Dr. Herman Goldenberg became the chief of the service. Goldenberg had trained in Munich and did postgraduate work for several years in Europe. During his tenure as Chief of Dermatology, Mount Sinai was among the first to introduce modern methods in the treatment of syphilis (Fig. 7). He received some of the original allotment of salvarsan before the drug was released for general use. It was said that he treated patients with early syphilis with miraculous results (5).

As the years passed, the department of dermatology acquired eight beds with the privilege of a number of beds in the pediatric wards. These beds were always occupied. Numerous interesting cases were presented at the various grand rounds. Dr. Goldenberg appeared in the 1914 photograph of the "Mount Sinai Giants of Yesteryear." (Fig. 8).

When Dr. Goldenberg retired in 1922, Dr. Walter J. Highman became the Chief of Dermatology (Fig. 9). Highman was a 1905 graduate of the Columbia College of Physicians and Surgeons. He did 2 years of postgraduate work in Berlin, Paris, Freiberg, Vienna, and in Bern under Dr. Josef Jadassohn. It was during his years at Mount Sinai that Highman established the first dermatopathology section at Mount Sinai. The need for this arose as a result of the special characteristics of the pathology of the skin (6). Dr. Highman published over seventy articles and was the author of the textbook *Dermatology* published in 1921.

On the death of Dr. Hyman in 1933, Dr. Isidore Rosen became Chief of Dermatology. Rosen was born in New York and graduated from the College of Physicians and Surgeons in 1903. He did pioneering work in the standardization of modern treatments of acquired syphilis (7). He was the first in America to organize a clinic for the systematic treatment of congenital syphilis. He was also the first to apply fever therapy in the treatment of chronic, recurrent dermatoses and to make chemical and biologic studies of the blood and sweat of these patients (Fig. 10).

During Rosen's tenure, laboratory facilities were added for dermatopathology, for dark field examinations for the recognition of early syphilis, and for the microscopic examination of hairs and scales for the diagnosis of fungal diseases. The clinic for syphilis therapy and congenital syphilis had two full-time medical assistants administering treatments (5). Many skin diseases were treated by radiotherapy, which was administered by dermatologists. At the time, most dermatologists administered radiotherapy to the skin in their offices.

During the 1930s, a contribution made by staff members caused a major sensation; it was announced on April 12, 1940 (8). A 5-day treatment for syphilis was developed by Louis Chargin (Fig. 11), Harold T. Hyman, and William Leifer. Chargin first suggested that the slow-drip method might be used for the injection of large amounts of arsenicals in syphilitic patients. Patients received 4 g in 5 days of neoarsphenamine. This reduced the cost of treating syphilis to less than ninety dollars. For a time, it looked as if Hyman, Chargin, and Leifer had realized Erlich's hope for the "magic bullet." However, the course of treatment of syphilis was changed forever with the discovery of penicillin.

Chargin was born in Minsk in 1881, grew up in New York, and graduated the University of Maryland School of Medicine in 1902. He did postgraduate study in Berlin and London from 1902 through 1904. He maintained a long association with Mount Sinai. He was also a skilled practitioner and maintained a prestigious practice on Central Park West. Chargin wore white gloves, bedroom slippers, and a visor in his office. He would never touch the fees he received. Patients were to slip the fees through a slit in a shoe box fashioned by Chargin. On his death, Chargin bequeathed the Chargin Fund for Dermatologic Research in the department.

On Rosen's retirement in 1946, Dr. Samuel Peck assumed the position as Director of Dermatology, a post he held until 1971. Peck was born in New York City in 1900 and graduated from the Columbia College of Physicians and Surgeons in 1924. He trained in New York City under Fordyce, and in Berlin and Zurich under Professor Bruno Bloch (Fig. 12). He made exceptional contributions in the field of contact and industrial dermatoses, insect-bite reactions, melanin, and mycology. His article, "Introduction of the Use of Mocassin Venom for the Treatment of Bleeding Symptoms" caused a sensation in 1932 (9).

During Peck's years as chairman of the department, a residency program was established for the training of dermatologists. He delivered the first paper at the first meeting of the Society for Investigative Dermatology in 1938, which was entitled, "The Use of Undecilenic Acid in the Treatment of Tinea." He also maintained an active clinical practice with a massive following. His office was located on East 70th Street, and was crowded with patients from before five in the morning until the early afternoon (Fig. 13).

Peck published more than 100 articles and co-authored or authored three books, "Occupational Diseases of the Skin" with Louis Schwartz and Louis Tulipan, "Cosmetics and Dermatitis" with Louis Schwartz and Paul Hoeber, and "Therapy of Dermatologic Disorders" with George Klein. He continued his private practice until nearly his 90th year, when he retired. He attended grand rounds in the department for a few years after this, contributing actively to the clinical discussions.

On Peck's retirement in 1971, Arthur Glick was named Acting Chairman, a position he held for 4 years (Fig. 14). Glick published several papers on industrial dermatoses and expanded the dermatologic surgery clinic. An avid fisherman, jazz enthusiast, and cigar aficionado, Glick brought in many new attendings and was the first to send residents in dermatology to the Bronx Veterans Hospital. Glick exerted a strong presence in the clinic with his sonorous voice and dapper bow-tie. He also maintained an elegant practice on East 67th Street on the top floor of a brownstone.

Another outstanding attending for many years was Orlando Cañizares (Fig. 15). He was born in Cuba, studied medicine in Paris and dermatology in Madrid, but he settled in New York after World War II. It was said that his life was influenced by three dictators: Cuban, Spanish, and German (10). His most important writings were on tropical dermatology, venereal diseases, the treatment of parapsoriasis with vitamin D2, and atrophoderma of Pasini and Pierini. He was regarded as dermatology's goodwill ambassador to tropical countries. He published several volumes of *Tropical Dermatology,* which became a classic. He was also an accomplished artist and found time to draw caricatures of many dermatologists.

On Glick's retirement in 1975, Harry Shatin became Chairman of the Department. He placed a special emphasis on the teaching of residents and the department excelled in clinical dermatology. Shatin was the first full-time chairman of the department (Fig. 16). A noted venerealogist, Shatin was an astute clinician. On one occasion, Shatin presented a case of bullous pemphigoid to his residents. He told his orderly of the diagnosis before the conference. During the conference, when none of the residents was able to diagnose the patient's condition with certainty, Shatin asked the orderly who promptly proclaimed, "Bullous pemphigoid!" At this, Shatin exclaimed, "You see, even the orderly knows the diagnosis." Shatin was credited with the organization and recredentialing of the residency program.

On Harry Shatin's retirement in 1979, Dr. Raul Fleischmajer was appointed chairman of the department. Fleischmajer was the first research-oriented chairman, having contributed widely to the areas of research in collagen vascular diseases, scleroderma, and dislipidoses. He was the first to measure collagen in a quantitative fashion in the skin. He was also instrumental in designating scleroderma a disease of inflammatory origin and in documenting the association of scleredema and diabetes mellitus. In his years at Mount Sinai, he was widely involved in the study of the nature of collagen fibrillogenesis.

Fleischmajer was also an astute clinician and maintained an active, hospital-based practice. He was the most widely travelled of all of the chairmen and had a worldwide reputation (Fig. 17). His lectures were most compelling from the scientific standpoint and also quite entertaining. He would share with residents and attendings alike his research ventures in places like Reims, which is, by no coincidence, the champagne capital of the world. Fleischmajer was also a connoisseur of opera and fine food.

During the Fleischmajer years, several prominent dermatologists joined the full-time staff. Dr. Irwin Kantor, a long-time associate of Dr. Peck, added savvy and experience to the department. He is known for his astute diagnoses of diseases that would baffle even the keenest of clinicians. Dr. Robert Phelps became the first full-time dermatopathologist and greatly expanded the division of dermatopathology. The department reads more than 50,000 slides yearly. Dr. Leonard Kornblee, a long-time Mount Sinai attending, became the full-time Chief of Dermatology at the Bronx Veterans Hospital. Dr. Marsha Gordon, a fine teacher and clinician, joined the department, as well as Dr. Donald Rudikoff, who is in charge of the Dermatology-HIV clinic. Dr. Ronald Shelton joined the department as its first Moh's chemosurgeon. Dr. Mark Lebwohl is currently the Director of the Division of Clinical Dermatology and Chairman. He has contributed widely in the field of psoriasis and pseudoxanthoma elasticum.

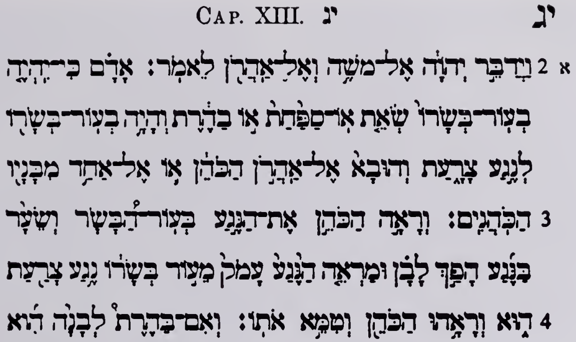
The department is currently involved in the evaluation of aging of the skin, the molecular mechanisms in ultraviolet light skin aging, and in pseudoxanthoma elasticum as a model for accelerated aging. It has been involved in the classification of the structure and function of connective tissue. Numerous clinical trials are underway in the areas of psoriasis, acne, atopic dermatitis, herpes simplex, and fungi.

The current voluntary attending staff has contributed greatly to the teaching program and to the field of dermatology. Dr. Robert Berger, Dr. Hillard Pearlstein, and Dr. Douglas Altchek are all notable dermatologic surgeons who have disseminated their knowledge of hair transplantation and dermabrasion. Dr. Stephen Kurtin, whose father pioneered dermabrasion, and Dr. David Orentreich, whose father pioneered hair transplantation, have both continued in the tradition. Dr. Ronald Sherman has been an innovator in combining the fields of cosmetology and dermatology. Dr. Michael Kalman has been creative in dermatologic surgery and dermatopharmacology. Dr. Albert Lefkovits, a fine clinician and teacher, is quite active in hospital affairs. Dr. Michael Fellner, an expert in immunodermatology, has helped bridge the gap in our understanding of the immunology of the skin. Many others, who are products of the dermatology program at Mount Sinai, have continued on the staff as fine teachers: Dr. George Jordan, Dr. Ellen Blank, Dr. David Cohen, Dr. Robert Warner, Dr. Susan Bershad, and Dr. Peter Wisch.

In this age of technology, the practice of dermatology continues to expand in many exciting directions. The Department of Dermatology is meeting the challenges of the future as innovative researchers and clinicians while maintaining the meticulous traditions of the past.

References

1. The Pentateuch and Haftorah, London, 1961.
2. Hirsh J, Doherty B. The first hundred years of the Mount Sinai Hospital. New York, 1952.
3. Fox GH. Photographic illustrations of skin diseases. New York, 1881.
4. Interview with Dr. Richard Hoffman, May 5, 1939.
5. Interview with Dr. Isidore Rosen. January 10, 1939.
6. Interview with Dr. Samuel Peck. April 16, 1951.
7. Centennial History of the American Dermatological Association. Chicago, 1976.
8. The five day treatment of syphilis by means of the continuous slow intravenous drip method. Presented to the New York Academy of Medicine, April 12, 1940.
9. Introduction of the use of moccasin venom for the treatment of bleeding symptoms. JAMA 1932.
10. Fisher A. Orlando Cañizares, A Eulogy. Presented at the Greater New York Dermatological Society, April 12, 1993.



CAP. XIII. יג יג

א 2 וַיְדַבֵּר יְהוָה אֶל־מֹשֶׁה וְאֶל־אַהֲרֹן לֵאמֹר׃ אָדָם כִּי־יִהְיֶה
בְעוֹר־בְּשָׂרוֹ שְׂאֵת אוֹ־סַפַּחַת אוֹ בַהֶרֶת וְהָיָה בְעוֹר־בְּשָׂרוֹ
לְנֶגַע צָרָעַת וְהוּבָא אֶל־אַהֲרֹן הַכֹּהֵן אוֹ אֶל־אַחַד מִבָּנָיו
3 הַכֹּהֲנִים׃ וְרָאָה הַכֹּהֵן אֶת־הַנֶּגַע בְּעוֹר־הַבָּשָׂר וְשֵׂעָר
בַּנֶּגַע הָפַךְ לָבָן וּמַרְאֵה הַנֶּגַע עָמֹק מֵעוֹר בְּשָׂרוֹ נֶגַע צָרַעַת
4 הוּא וְרָאָהוּ הַכֹּהֵן וְטִמֵּא אֹתוֹ׃ וְאִם־בַּהֶרֶת לְבָנָה הִוא

Fig. 1. Leviticus, Chapter XIII, Verse 3.

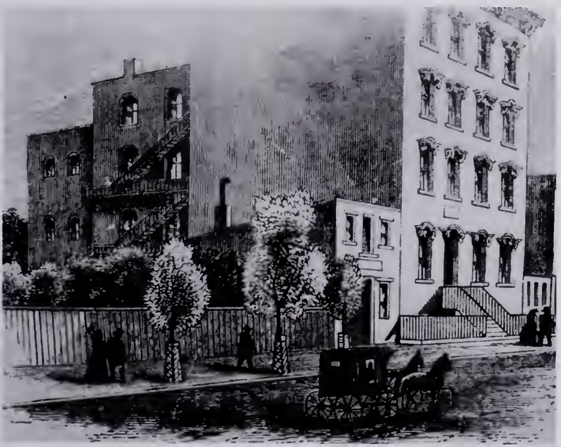


Fig. 2. The first Mount Sinai Hospital, located at 28th Street between 7th and 8th Avenues.

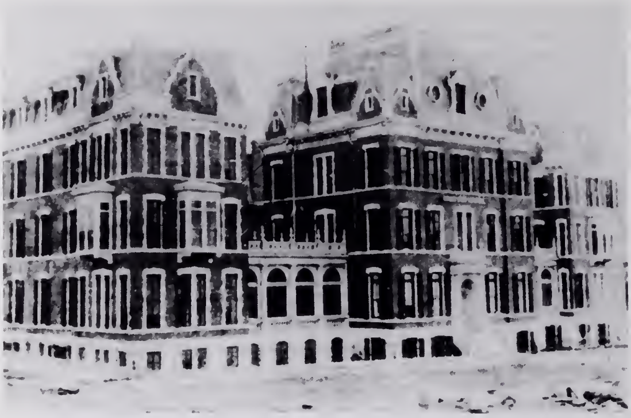


Fig. 3. The second Mount Sinai Hospital at Lexington Avenue between 67th and 68th Streets.

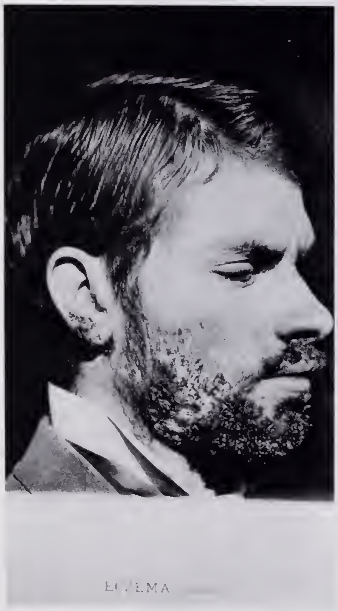


Fig. 4. Eczema circa 1881.

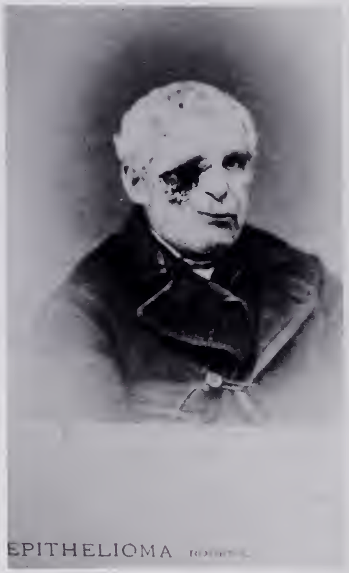


Fig. 5. Basal cell carcinoma circa 1881.

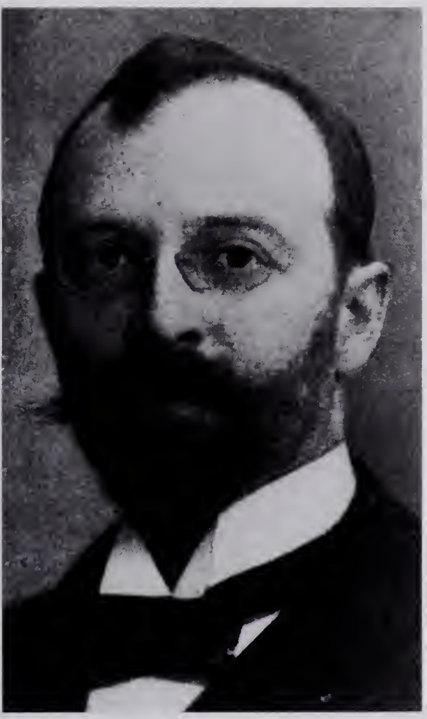


Fig. 6. Dr. Sigmund Lustgarten, Chief of Dermatology, 1891–1910.

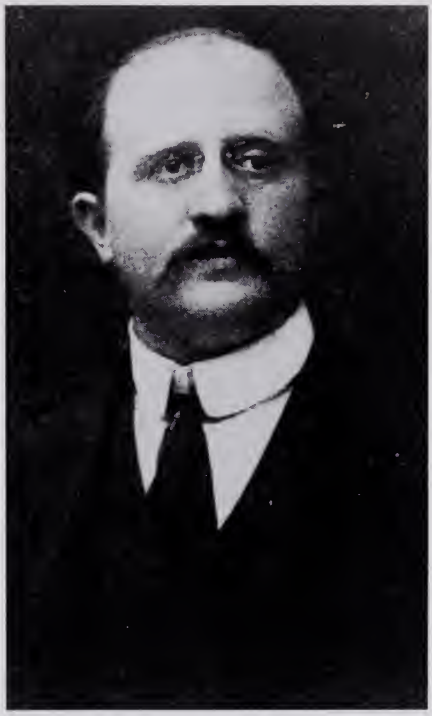


Fig. 7. Dr. Herman Goldenberg, Chief of Dermatology, 1910–1922.

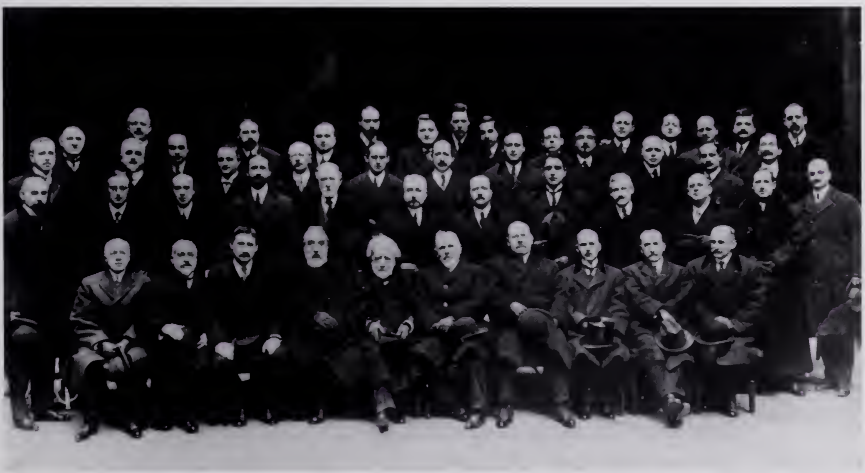


Fig. 8. The Mount Sinai Giants of Yesteryear, 1914. Dr. Goldenberg is in the second row, seventh from left to right.

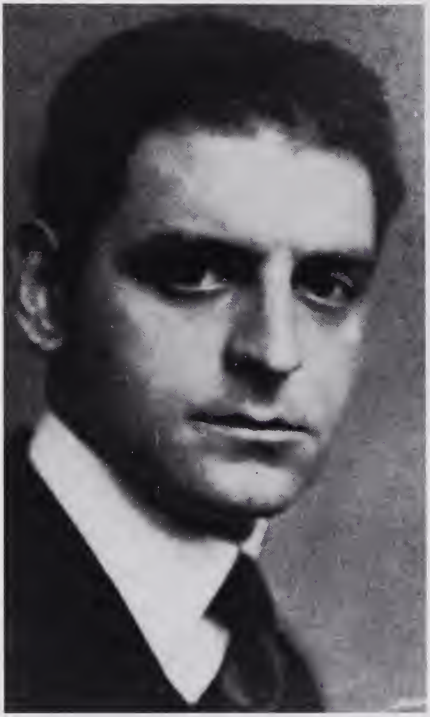


Fig. 9. Dr. Walter Highman, Chief of Dermatology, 1922–1933.

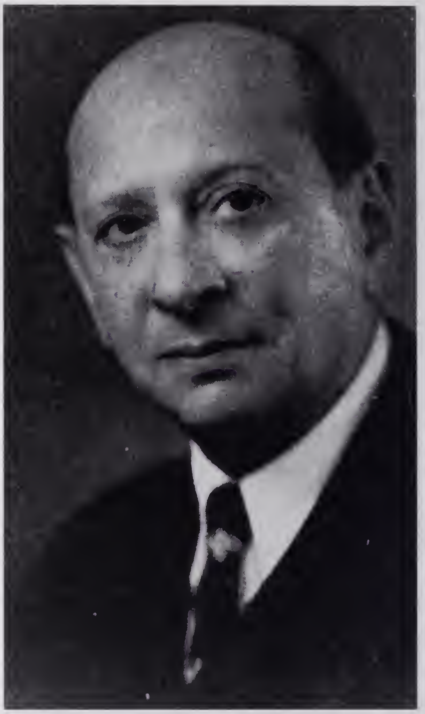


Fig. 10. Dr. Isidore Rosen, Chief of Dermatology, 1933–1945.

LEVY LIBRARY—MOUNT SINAI

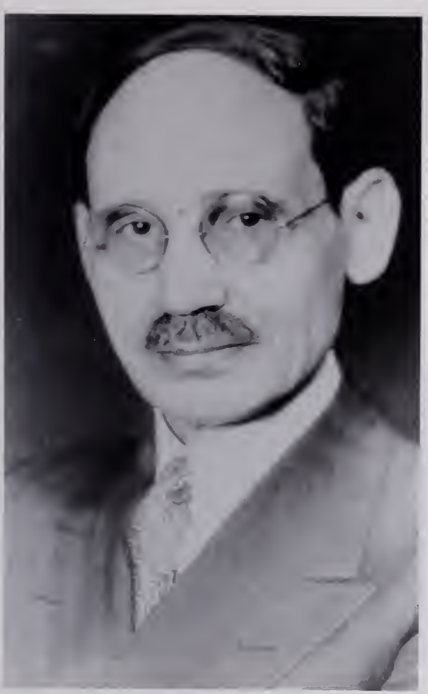


Fig. 11. Dr. Louis Chargin, 1952.

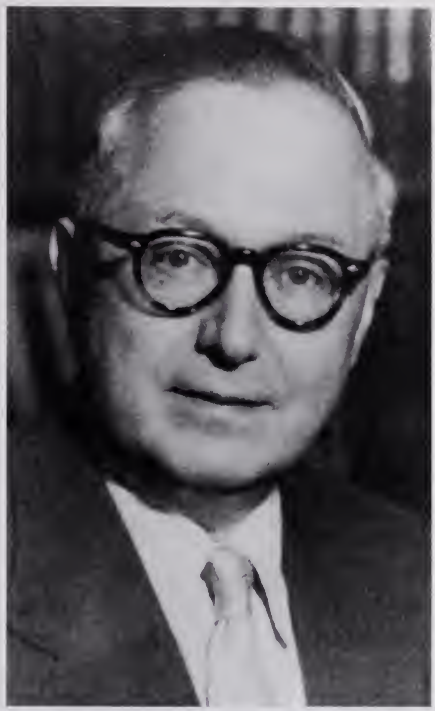


Fig. 13. Dr. Samuel Peck, Chief of Dermatology, 1946–1971.

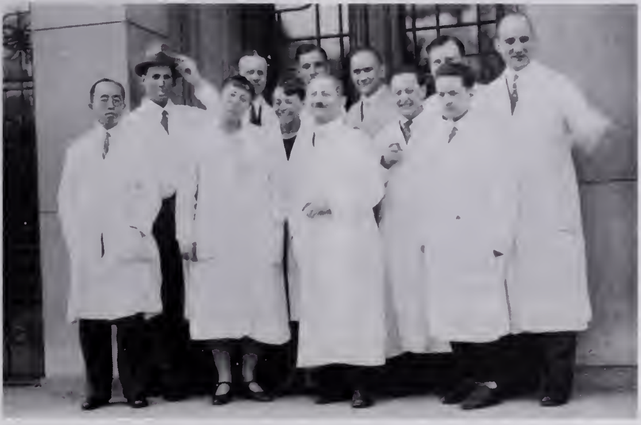


Fig. 12. Dr. Samuel Peck, front row to the right of Prof. Bloch (center), Zurich, 1927.

Fig. 14. Dr. Arthur Glick, Chief of Dermatology, 1971–1975.

Fig. 15. Dr. Orlando Cañizares, 1976.

Fig. 16. Dr. Harry Shatin, Chief of Dermatology, 1975–1979.

Fig. 17. Fifty years of Dermatology Chairmen: left to right Fleischmajer, Shatin, Peck, and Glick.

LEVY LIBRARY—MOUNT SINAI

Vena Cava Filters:

Available Devices and Current Practices

JAMES M. COOPER, M.D., JAMES SILBERZWEIG, M.D., AND HAROLD A. MITTY, M.D.

Abstract

Deep venous thrombosis and pulmonary embolus are significant clinical problems. Although anticoagulation remains first-line therapy for thromboembolic disease, the placement of inferior vena cava filters plays an important role as alternative or supplemental therapy. Initial filters were high-profile devices intended to be placed via surgical cutdown. Concerns about the size of the delivery systems prompted development of low-profile systems intended specifically for percutaneous placement. The placement of filters has progressively become a percutaneous procedure performed by interventional radiologists outside of the operating room. This article reviews the currently available devices that have been approved by the Food and Drug Administration, their indications, efficacy, and complications. Experimental temporary filter designs are discussed as well.

DEEP VENOUS THROMBOSIS (DVT) and pulmonary embolism (PE) are significant clinical problems. It has been estimated that there are at least 170,000 diagnosed initial episodes and 90,000 recurrent episodes of thromboembolism a year in the United States (1). Accounting for underdiagnosis, the true incidence may be as high as 600,000 cases a year (1, 2). It has been estimated that PE is directly responsible for approximately 100,000 deaths a year and is a contributing cause of death in an additional 100,000 patients in the United States (3, 4). A 30% incidence of DVT has been identified in general surgical postoperative patients who do not receive anticoagulant prophylaxis. Four percent will suffer symptomatic PE and 1 to 2% will suffer fatal PE (1). There is a 40 to 50% incidence of DVT in orthopedic surgical patients without prophylaxis and a 1 to 2% rate of fatal PE after total hip replacement (5). Subclinical emboli have been found in 64% of patients undergoing autopsy after dying of various causes (1).

Risk factors identified for DVT include: older age, immobilization, previous DVT, surgery, general anesthesia, pregnancy, malignancy, hypercoagulable states, and trauma. Traditionally, anticoagulation has been the initial treatment in patients with DVT or PE. A course of therapy usually includes 5 days of intravenous or subcutaneous heparin followed by oral warfarin for 6 weeks to 6 months (2, 6). In most cases, longer courses of therapy have not been shown to improve outcome. Nevertheless, longer regimens can be administered in patients with prolonged immobilization, persistent hypercoagulable state, or recurrent DVT. The incidence of hemorrhagic complications of anticoagulation requiring transfusion ranges from 1.5 to 20% (4, 6). Overall mortality rates due to anticoagulation are reported at 5 to 12% (6).

Although anticoagulation remains first-line treatment for DVT/PE, vena cava filters are indicated in patients with a contraindication to anticoagulation, in patients with recurrent PE despite therapeutic anticoagulation, and in patients with complications from anticoagulation.

Before the availability of inferior vena cava (IVC) filters, surgical interruption of the IVC was performed to prevent PE. A variety of direct sur-

---

From the Department of Radiology, Mount Sinai Medical Center, New York, NY. Address reprint requests to James M. Cooper, MD, Assistant Professor, Radiology, Department of Radiology, Mt. Sinai Medical Center, Box 1234, One Gustave L. Levy Place, New York, NY 10029.

gical procedures including ligation, plication, and placement of extracaval serrated clips were performed and have all but been abandoned (2, 6). The attendant risks and complications of such surgical procedures prompted the development of transvenous interruption devices. In 1967, the Mobbin-Uddin umbrella was introduced and was the first widely available device. The umbrella was introduced through a surgical cutdown of the right internal jugular vein. The device was abandoned eventually after caval thrombosis rates of 60 to 73% were reported (2, 6, 7). In the early 1970s, the stainless steel Kimray-Greenfield (KG) filter (Medi-tech, Watertown, MA) was introduced. This device was initially designed to be placed via a surgical cutdown of either the right internal jugular vein, the right common femoral vein, or the right atrium (8, 9). Despite the fact that the introducer system consists of a 29 French sheath and 24 French delivery system, percutaneous placement of this device, primarily by interventional radiologists, became common practice during the 1980s (6, 7, 10–14). Concerns about the size of the delivery system and high reported rates of access site thrombosis of 10 to 24% (7, 11, 15) after percutaneous placement prompted development of lower profile devices designed specifically for percutaneous placement (16–21). Currently, in addition to the KG stainless steel filter, five lower profile IVC filters have been approved by the Food and Drug Administration and are available for use in the United States. As a result, the placement of caval filters has progressively become a percutaneous procedure performed by interventional radiologists outside of the operating room (2, 6, 22–26).

Indications

Because vena cava filters can be placed easily without the need for cutdown, there has been some debate over indications for these devices. Traditional indications for IVC filter placement include: documented PE or DVT with a contraindication to anticoagulation, failure of anticoagulation, complication of anticoagulation, or free-floating IVC thrombus even in patients whose blood can be anticoagulated. More recently, advocated uses include placement in patients with DVT or PE who have poor pulmonary reserve and may not tolerate repeated emboli and in patients who have suffered a large PE and who might not survive an additional insult (27). Although some debate exists over what constitutes a contraindication to anticoagulation, the indications above are generally accepted by most clinicians (4). More controversial are the proposals for prophylactic placement of filters in asymptomatic patients who may be at increased risk for DVT/PE. These high-risk patients include those who have suffered trauma, who have undergone orthopedic surgery, and who are bedridden (27). Equally controversial is the proposed use of filters as first-line therapy in oncology and obstetric patients with DVT (23, 25, 28, 29). Enthusiasm for embracing liberalized indications for filter placement varies from institution to institution and from indication to indication. In our practice at The Mount Sinai Medical Center, we do not place prophylactic filters routinely. A contraindication to anticoagulation in patients with DVT or PE is the most common indication for filter placement.

It is clear that the number of filters placed nationally has increased since the advent of percutaneous devices (22–24, 26). One institution found an increase in utilization after the Interventional Radiology Service began to place filters; although no liberalization of indications was identified. The increased utilization was attributable to placement in critically or terminally ill patients who would not have been considered surgical candidates (23, 26). Despite the dismal prognosis of patients with illnesses such as advanced malignancy, filter placement has been advocated because anticoagulation is often contraindicated (23, 25, 29). Although mortality rates at 6 months may be high secondary to underlying disease processes, those patients who do survive seem to benefit for the remainder of their lives by having fewer episodes of PE (23).

Available Devices

The original KG stainless steel filter is a conical device designed to maintain caval flow despite progressive occlusion of the filter by thrombus (2, 15). Placement of the device is limited to IVCs measuring no greater than 28 mm in diameter. The device is placed through a 29 French sheath. There is a reported long-term caval patency rate of 95% and a recurrent embolism rate of 4% (8). These rates traditionally have been considered the "gold-standard" to which other filter designs must be held. However, the validity of these data has been questioned because of the subjective, mostly clinical, nature of the follow-up studies. In fact, data from the Mayo Clinic suggest a much higher and unsuspected caval thrombosis rate of 29% in patients studied more objectively with computerized tomography (30). Tilting of the device more than 15% off its axis has been associated with decreased clot-trapping capability (2, 15). Multiple case reports of caval perforation by filter struts have been published with penetration into a number of structures including the ureter, aorta, intestine, and vertebral body. Strut perforation is reported to occur in up to 41% of cases; however, the overall rate of symptomatic perforation is thought to be very low (6, 8, 15, 31).

Malpositioning of the filter at the time of placement has been described. In surgical series, initial malposition of the KG filter was reported in 4 to 9% of cases (6, 8, 32). Migration of the device (Fig. 1) has been reported (5, 10, 33). Insertion site thrombosis after percutaneous placement of the 29 French system was significant with an incidence of 10 to 24% (2, 6, 11–13).

The titanium version of the KG filter (Meditech) became available in the late 1980s and was originally designed to be delivered via a 14 French sheath. The filter also has a conical design to maximize blood filtration even when thrombus is ensnared in the filter. Titanium is more elastic than stainless steel and allows for greater compression of the device into the smaller delivery system. Initial design of the filter legs and hooks was later modified to enhance attachment to the caval wall and decrease the initial reported high incidence of perforation and migration (2). The newer modified hook version appears to have a lower penetration rate (6, 7). In addition, to facilitate passage of the relatively rigid filter from a femoral approach, the device is now supplied with a reinforced 15 French sheath. Studies of the titanium Greenfield filter (TGF) revealed recurrent PE rates of 3 to 4% and caval patency rates of approximately 91 to 100% (2, 3, 6, 7).

As expected with the smaller delivery system, the incidence of insertion vein thrombosis is lower than that of the original KG filter (2, 12, 34). The TGF, unlike the original KG filter, is loaded in a self-contained capsule and deployed by a one-step mechanical release mechanism. Despite the relative ease of deployment, problems have been identified. Studies have shown asymmetry of the legs occurs in 2 to 71% of cases (2, 34). True interlocking of the legs has been reported as well. *In vitro* and *in vivo* studies have suggested but not proven that such asymmetry and interlocking reduces clot-trapping efficiency (2, 30, 31, 34, 35). Attempts at guidewire or catheter manipulation have been successful in correcting initial deployment asymmetry, but there is an attendant risk of filter migration (34, 36, 37). Tilting has also been reported in 23% of the cases and has been attributed to the lack of a central guidewire to help center the filter in the IVC (34, 35). As with all filter designs, initial malpositioning of the filter can occur, usually secondary to technical error (38). The titanium filter is less ferromagnetic than the stainless steel version and, therefore, exhibits less artifact on magnetic resonance imaging studies.

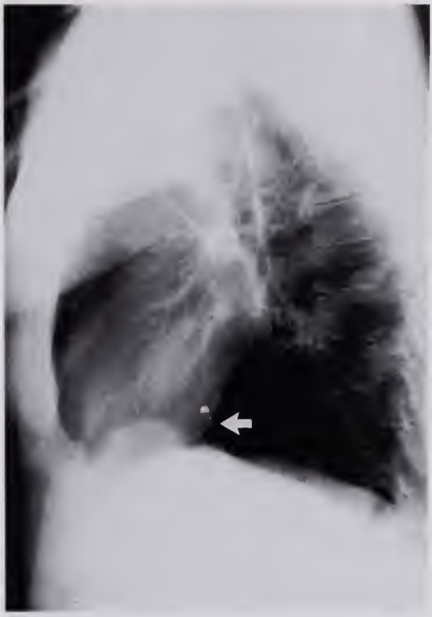


Fig. 1. Stainless steel Greenfield filter projecting into the right atrium (arrow).

The newest version of the Greenfield filter is a smaller profile stainless steel version (STG) (Medi-tech) (Fig. 2). This design became available in 1995 and is delivered through the same 15 French introducer sheath as the TGF. The actual deployment device is similar to that of the titanium filter with an important exception: the newer filter is placed over a wire. This is thought to provide better initial centering and to facilitate retrieval in the event of malpositioning. An *in vitro* study showed no significant difference in clot-capturing ability of this filter relative to the two other Greenfield filter versions (37). A recent clinical study reported placement in 75 patients. After 30 days follow-up, a 95% caval patency rate was reported. There was no clinically suspected PE and one case of caval penetration. Interestingly, there was a 5% incidence of initial asymmetry (39). Large clinical series using this device have not yet been published.

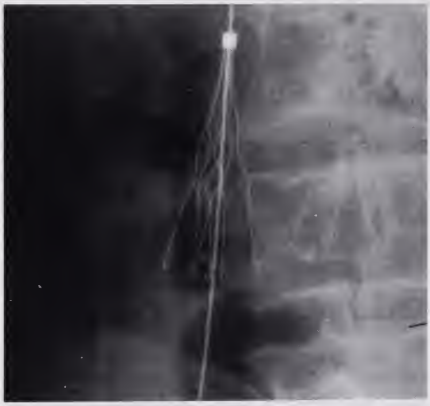


Fig. 2. New low-profile over-the-wire stainless steel Greenfield filter.

The Vena-Tech filter (VTF) (B. Braun Medical, Evanston, IL) (Fig. 3) is a modified cone device that comes prepackaged in a 10 French carrier and requires a 12 French delivery sheath. It became available in the United States in 1986. The six side rails of the device were designed to reduce the incidence of tilting. The filter was also designed only for IVCs up to 28 mm in diameter. Rates of caval patency range from 76 to 96%, and recurrent PE rates range from 2 to 3% (2, 6, 7, 15, 19, 20). *In vivo* and *in vitro* studies suggest a greater clot-capturing efficiency than the TGF (40). Despite the reported advantages of the stabilizing side bars, tilting and migration have been reported, although at lower rates than those reported for the Greenfield devices (2, 6, 7, 15, 19, 20, 41). The actual deployment of the device is more operator dependent than that of the low-profile Greenfield devices. Failure to release the filter in one rapid, smooth motion has been associated with increased tilting (6). It has been postulated that the incidence of tilting decreases with operator experience (15).

The Simon nitinol filter (SNF) (Bard Radiology, Covington, GA) (Fig. 4) is made of an alloy of nickel and titanium, nitinol, that has a unique thermal memory property. When cooled, the alloy becomes very flexible. When warmed to body temperature, the device assumes a preformed shape (15). The filter consists of a dome with six fixation legs. The delivery system is the lowest profile of all the available devices, with a 7 French carrier system and a 9 French sheath. Recurrent PE rates of approximately 1% have been reported (6, 15). *In vitro* and *in vivo* studies have shown that the clot-trapping efficiency of the SNF is the greatest of all the available devices (42, 43). The reported caval patency rates range from 75 to 90% (2, 7, 15). Migration of the device has been reported in 2 to 3% of cases (2, 44). The filter can only be used in IVCs up to 28 mm in diameter; it is also nonferromagnetic. The marked flexibility of the system allows deployment from an antecubital approach as well as relatively smooth deployment from the left common femoral vein. Use of an antecubital vein can facilitate placement in an ambulatory setting. In addition, access vein thrombosis in the arm is relatively unimportant. The filter must be cooled with iced normal saline

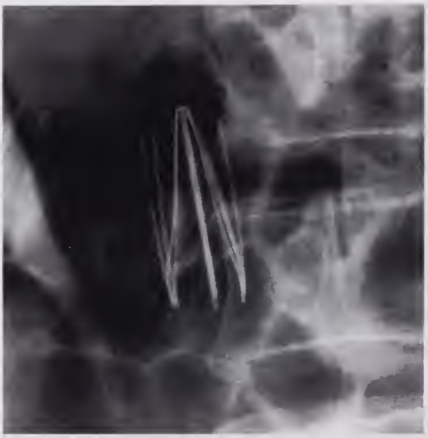


Fig. 3. Vena-Tech filter.

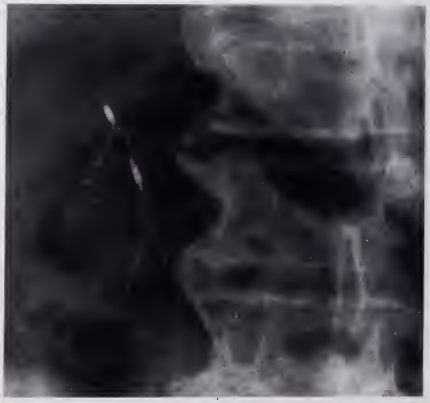


Fig. 4. Simon nitinol filter.

before deployment and assumes its shape inside the body when released.

The Bird's Nest filter (BNF) (Cook, Bloomington, IN) (Fig. 5) was the first low-profile device to be approved for use in the United States. It consists of stainless steel wires shaped into a criss-crossing pattern that assumes the shape of a bird's nest in the IVC. The wires are secured by four hook wire struts. The filter comes loaded in a 12 French delivery system, requiring a 14 French sheath. The rate of recurrent PE is reported at 0.5 to 4% (2, 6, 7, 16). Although the caval occlusion rate associated with the BNF is reported traditionally at approximately 3% (6, 7, 16), a more recent study reported a rate of 15% (2). No migrations have been reported (7, 45). The distinct advantage of the BNF is the ability to place the device in IVCs up to 40 mm in diameter (15). However, the device is the most operator dependent of all available filters and is considered by some to be the most technically challenging (7, 46). The overall cranio-caudal extension of the filter ranges from 3.5 to 13.6 cm depending on the deployment technique (45). The relative long length of the device can limit its use in short infrarenal IVCs or those with infrarenal clot (46). Prolapse of filter wires above the wire struts has been described, but this is not associated with decreased filtering efficiency (2, 45, 46).

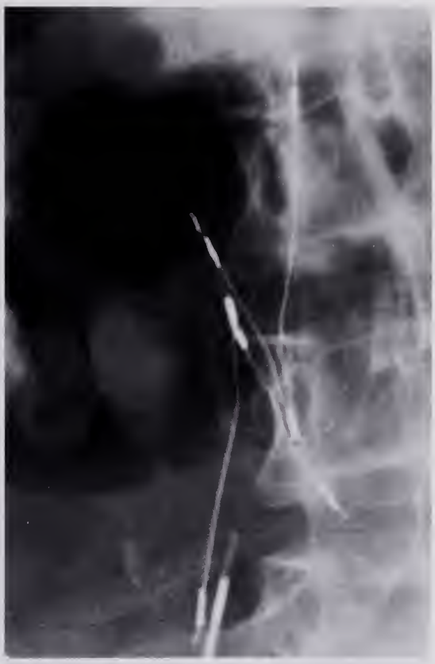


Fig. 5. Bird's Nest filter.

Choosing a Device

The characteristics of an ideal filter would include: clot-trapping efficiency of both large and small emboli without compromise of caval flow/patency; secure attachment to the caval wall; a nonthrombogenic, biocompatible, and long-lasting material; a low-profile introducer system; a nonferromagnetic composition; low cost and retrievability (6, 15). No one filter currently available meets all of the criteria. None of the available filters approved by the Food and Drug Administration are retrievable.

The choice of a device in a given patient depends on the specific clinical situation and the experience of the operator with the available devices. For example, in patients with a long life expectancy, concerns about filtering efficiency and caval patency are paramount (6). In other patients, risk of periprocedure bleeding may be a significant concern, necessitating the lowest profile introducer system. Ease of insertion becomes a concern when traditional approaches such as the right common femoral vein or right internal jugular vein are unavailable. More flexible, lower profile systems are better suited to other less direct routes of venous access, such as the antecubital vein or left common femoral vein. In addition, caval size is a determining factor in filter selection. Only the BNF is approved for use in IVCs larger than 28 mm in diameter.

Complications

Regardless of the device selected, certain periprocedural and long-term complications of filters have been identified (6, 7). When filters are placed by cutdown, the incidence of major bleeding requiring re-exploration is 6 to 15%. When placed percutaneously, there is a 10% incidence of access vein thrombosis even with the low-profile devices (11). Inadvertent deployment into the right atrium (RA), renal veins, hepatic veins, and iliac veins has been described. Initial malpositioning had been attributed to technical error and lack of a quality inferior vena cavagram (6, 8, 32). Tilting of the device at placement has been described. Other potential, more long-term problems are caval perforation and filter migration. The incidence of caval thrombosis and recurrent

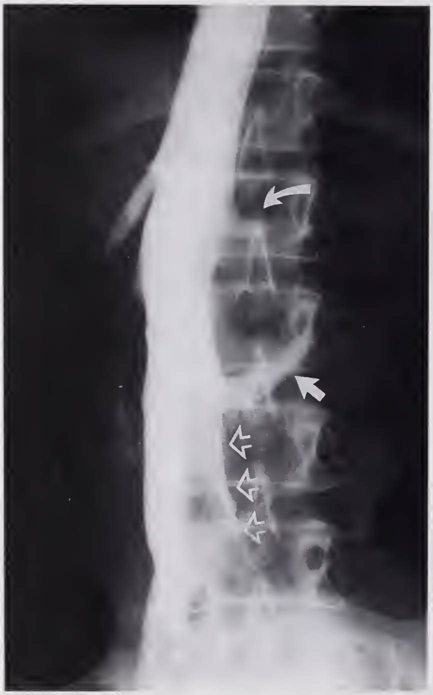


Fig. 6. Inferior vena cavagram revealing intraluminal thrombus (open arrows) extending up to retroaortic portion of the circumaortic left renal vein (arrow). The more cephalad anterior limb is identified (curved arrow). The retroaortic limb may act as a collateral pathway for clot around a filter, necessitating a suprarenal placement in this patient. In a patient without IVC thrombus, a filter must be placed below the retroaortic limb to provide adequate protection.

PE after placement are recognized as the best bench marks of long-term efficiency and effectiveness (6). Phlegmasia cerulea dolens has been reported after placement of both the KG filters and the BNF (4, 47).

Temporary Devices

A major disadvantage of the devices above is that they are permanent. Currently, much effort is being channeled into the development of a temporary filter device. Such a device would be ideally suited to those patients in whom a filter is placed as a prophylactic measure or in whom there is a short-term contraindication to anticoagulation therapy (48). Three basic design types have been investigated. Detachable devices that are deployed and later retrieved, nondetachable devices that are tethered to the skin or subcutaneous tissue, and catheter type devices that can be used to facilitate mechanical venous thrombectomy or thrombolysis. Unlike the first two, the latter design is meant for very short-term placement during catheter-directed therapy. Such a filter is to be removed at the termination of the therapy (49). The others are designed to stay *in vivo* for a temporary period of days to weeks.

Different versions of a tethered nitinol wire filter have been studied (48, 50). These filters are essentially long nitinol wires with a spiral configuration at their distal ends. The proximal straight portion of the wire is tethered in place in the femoral vein with a vascular sheath. Perisheath thrombosis is a limitation of these devices (48, 50).

A completely subcutaneous tethered device has undergone clinical trials in Europe and animal trials in the United States (51). The Tempofilter (B. Braun Celsa, Chasseneuil, France) is made of phynox (42% cobalt, 21.5% chrome, 18% nickel, 8.84% iron, 7.5% molybdeum, 2% manganese, 0.15% carbon, 0.001% beryllium), the same alloy used in the VTF (51). It is designed to be introduced via the right internal jugular vein and held in place in the IVC by a 7 French shock-absorbing, nylon tethering device. The catheter is anchored subcutaneously at the venotomy site. The filter is conical but lacks fixation struts to facilitate removal. In a clinical study of 64 patients, retrieval was possible in 100% of attempted cases at up to 6 weeks. IVC occlusion occurred in 3% of patients and perifilter thrombus was noted in 14%. Recurrent PE was suspected, but not documented, in one patient (51). Migration to the right atrium occurred in two patients (51). As with all temporary designs, the accumulation of perifilter thrombus is problematic. Because such thrombus could conceivably break free at the time of filter removal, retrieval may require lytic therapy or placement of a permanent device (6, 52).

Detachable devices have been investigated as well. The Tulip-Gunther filter (Cook Europe, Bjaeverskov, Denmark) is a stainless steel device available in Europe and has a hook at the apex to facilitate retrieval. The hook is designed to be snared percutaneously (52). However, the retrieval period is limited to 2 to 3 weeks because of neointimal covering of the filter struts after this period (53). Newer devices are currently being evaluated that reportedly induce little or no neointimal reaction and can be removed safely after 6 weeks (53).

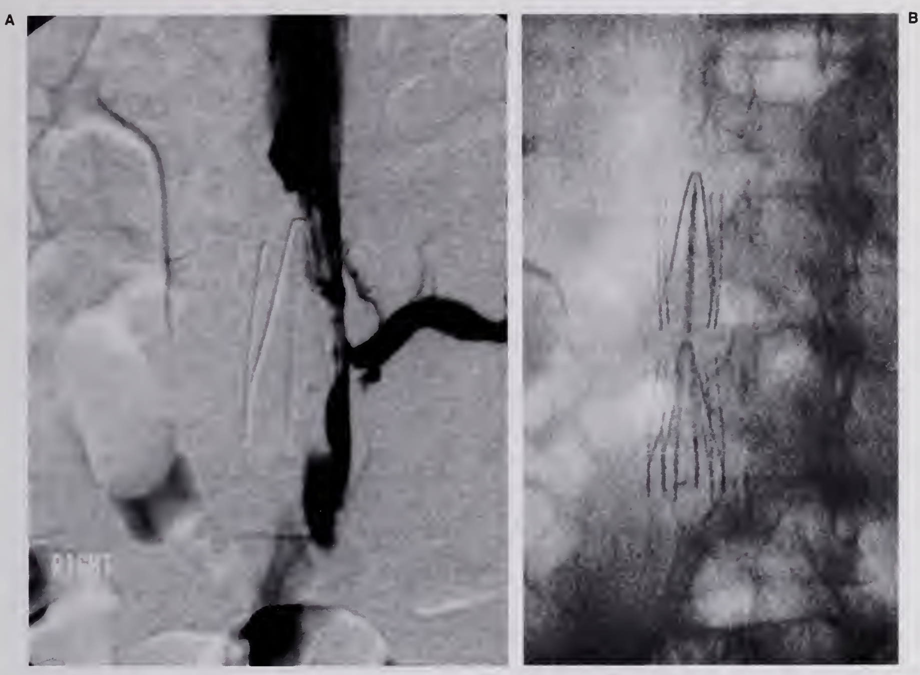


Fig. 7. A, Inferior vena cavagram revealing perifilter thrombus extending cephalad to a Vena-Tech filter in a patient with documented recurrent pulmonary embolus. Note the initial positioning across the orifice of the left renal vein necessitated by infrarenal IVC thrombus at the time of initial placement. B, After placement of a second, more cephalad filter.

The Role of the Interventional Radiologist

In most major medical centers, interventional radiologists place the majority of vena cava filters. IVC filtration has evolved into a therapy that can be optimize by using a variety of devices and skills used by the vascular radiologist. All of the new low-profile devices are placed in part using catheter and guidewire techniques. Filter placement from either a right internal jugular vein or right common femoral vein provides relatively straightline access to the IVC. Less direct routes such as the left common femoral vein, left jugular vein, or antecubital veins require more sophisticated manipulation. The importance of quality inferior vena cavagrams is well recognized in the placement of filters (6, 8, 54). An operator needs to assess caval diameter, carefully locate the renal veins, delineate the extent of thrombus within the IVC, and recognize any vascular anomalies that may have an impact on filter placement (6). Ideally, a filter should be deployed below the renal veins except when the presence of clot precludes such a location. Although the entrance of the renal veins into the IVC is usually found at L1–L2, a recent study by Hicks et al. (54) revealed a 12% incidence of locations above and below this level, from T12 to L4. In addition, anomalies of the renal veins or IVC occurred in 37% of 108 patients. These anomalies included circumaortic renal veins, multiple renal veins, retroaortic renal veins, and duplicated IVC (Fig. 6). They determined that in 18% of cases, these anomalies altered the position selected for the filter deployment (54). Clearly, adequate interpretation of the cavagram requires high-quality imaging. Quality imaging and sophisticated catheter manipulation are necessary not only in the initial placement of the devices but in treatment of complications as well. Percutaneous retrieval of initially malpositioned permanent filters is possible. In cases of recurrent PE or perifilter thrombus formation, placement of a second filter may be required. The precise delineation of the extent of thrombus and position of the second filter is extremely important (Fig. 7). The percutaneous placement and retrieval of temporary filters require advanced guidewire and catheter techniques as well as high grade imaging. In addition, there is a need for careful cavography to identify perifilter thrombus which may require lytic therapy before removing these filters. For these reasons, it is not surprising that the interventional radiologist plays the primary role in the placement of such devices.

Summary

Although none of the available vena cava filters meet the definition of the "ideal filter," the variety available allows the interventional radiologist to tailor his choice to a particular patient's needs and anatomy. Factors such as available access site, life expectancy, coagulation status, and caval diameter must be considered before treatment. Because there is no multicenter, controlled long-term prospective study comparing filters, the choice depends mostly on the physician's preference and experience with the devices.

References

1. Weinmann EE, Salzman EW. Deep-vein thrombosis. N Engl J Med 1994; 331:1630–1641.
2. Mohan CR, Hoballah JJ, Sharp WJ, et al. Comparative efficacy and complications of vena caval filters. J Vasc Surg 1995; 21:235–246.
3. Paiement GD. Thromboembolic complications following trauma surgery: Incidence and outcomes. Orthopedics 1995; 18(suppl):12–14.
4. Harris JE, Kinney EV, Harris EJ, et al. Phlegmasia complicating prophylactic percutaneous inferior vena caval interruption: A word of caution. J Vasc Surg 1995; 22: 606–611.
5. Turpie AG. Deep vein thrombosis prophylaxis in the outpatient setting: Preventing complications following hospital discharge. Orthopedics 1995; 18(suppl):15–17.
6. Dorfman G. Percutaneous inferior vena caval filters. Radiology 1990; 174:987–992.
7. Ferris EJ, McCowan TC, Caver DK, et al. Percutaneous inferior vena cava filters: Follow-up of seven designs in 320 patients. Radiology 1993; 188:851–856.
8. Greenfield LJ, Michna BA. Twelve-year clinical experience with the Greenfield vena cava filter. Surgery 1988; 104:706–712.
9. Greenfield LJ, Zocco J, Wilk J, et al. Clinical experience with the Kimray-Greenfield vena caval filter. Ann Surg 1977; 185:692–698.
10. Tadavanthy SM, Castaneda-Zuniga W, Salomonowitz E, et al. Kimray-Greenfield vena cava filter: Percutaneous introduction. Radiology 1984; 151:525–526.
11. Mewissen MW, Erickson SJ, Foley WD, et al. Thrombosis at venous insertion sites after inferior vena caval filter placement. Radiology 1989;173:155–157.
12. Molgaard CP, Yucel EK, Geller SC, et al. Access-site thrombosis after placement of inferior vena cava filters with 12–14-F delivery systems. Radiology 1992; 185;257–261.
13. Dorfman GS, Cronan JJ, Paolella LP, et al. Intragenic changes at the venotomy site after percutaneous placement of the Greenfield filter. Radiology 1989; 173:159–162.
14. Pais SO, Tobin KD. Percutaneous insertion of the Greenfield filter. Am J Roentgenol 1989; 152:933–938.
15. Grassi CJ. Inferior vena cava filters: Analysis of five currently available devices. Am J Roentgenol 1991; 156: 813–821.
16. Roehm JOF, Johnsrude IS, Barth KH, et al. The bird's nest inferior vena cava filter: Progress report. Radiology 1988; 168:745–749.
17. Simon M, Athanasoulis CA, Kim D, et al. Simon nitinol inferior vena cava filter: Initial clinical experience. Radiology 1989; 172:99–103.
18. Ricco JB, Crochet D, Sebilotte P, et al. Percutaneous transvenous caval interruption with the LGM filter: Early results of a multicenter trial. Ann Vasc Surg 1988; 3:242–247.
19. Crochet DP, Stora O, Ferry D, et al. Vena Tech-LGM filter: Long-term results of a prospective study. Radiology 1993; 188:857–860.
20. Millward SF, Peterson RA, Moher D, et al. LGM (Vena Tech) vena caval filter: Experience at a single institution. J Vasc Interv Radiol 1994; 5:351–356.
21. Greenfield LJ, Cho KJ, Pais O, et al. Preliminary clinical experience with the titanium Greenfield vena caval filter. Arch Surg 1989; 124:657–659.
22. Athanasoulis CA. Complications of vena cava filters. Radiology 1993; 188:614–615.
23. Crystal KS, Kase DJ, Scher LA, et al. Utilization patterns with inferior vena cava filters: Surgical versus percutaneous placement. J Vasc Interv Radiol 1995; 6:443–448.
24. Alexander JJ, Yuhas JP, Piotrowski JJ. Is the increasing use of prophylactic percutaneous ICV filters justified? Am J Surg 1994; 168:102–106.
25. Lossef SV, Barth KH. Outcome of patients with advanced neoplastic disease receiving vena caval filters. J Vasc Interv Radiol 1995; 6:273–277.
26. Magnant JG, Walsh DB, Juravsky LI, et al. Current use of inferior vena cava filters. J Vasc Surg 1992; 16:701–706.
27. Rohrer MJ, Scheidler MG, Wheeler HB, et al. Extended indications for placement of an inferior vena cava filter. J Vasc Surg 1989; 10:44–50.
28. Cohen JR, Tenenbaum N, Citron M. Greenfield filter as primary therapy for deep venous thrombosis and/or pulmonary embolism in patients with cancer. Surgery 1991; 109:12–15.
29. McLoughlin RF, Sirkis H, So CB, et al. Severity of disease score as a predictor of mortality after caval filter insertion. J Vasc Interv Radiol 1995; 6:715–719.
30. Dorfman G. Risks and benefits of manipulation of the titanium Greenfield inferior vena cava filter after deployment: Filter facts and filter fantasies. J Vasc Interv Radiol 1993; 4:617–620.
31. Lok S, Adkins J, Asch M. Caval perforation by a Greenfield filter resulting in small bowel volvulus. J Vasc Interv Radiol 1996; 7:95–97.
32. Van Allan RJ, Hanks SE, Harrell DS, et al. Percutaneous retrieval of a misplaced titanium Greenfield filter. Cardiovasc Intervent Radiol 1994; 17:110–112.
33. Yakes WF. Percutaneous retrieval of Kimray-Greenfield

filter from right atrium and placement in inferior vena cava. Radiology 1988; 169:849–851.

34. Sweeney TH, Van Aman ME. Deployment problems with the titanium greenfield filter. J Vasc Interv Radiol 1993; 4:691–694.
35. Korbin CD, Van Allan RJ, Andrews RT, et al. Strut interlocking of titanium Greenfield vena cava filters and its effect on clot capturing: An *in vitro* study. Cardiovasc Intervent Radiol 1994; 17:204–206.
36. Moore BS, Valji K, Roberts AC, et al. Transcatheter manipulation of a symmetrically opened titanium Greenfield filters. J Vasc Interv Radiol 1993; 4:687–690.
37. Greenfield L, Proctor MC. Experimental embolic capture by asymmetric Greenfield filters. J Vasc Surg 1992; 16: 436–444.
38. Cynamon J, Bakal CW, Epstein SB, et al. Percutaneous removal of a titanium Greenfield filter. Am J Roentgenol 1992;159:777–778.
39. Cho KJ, Proctor MC, Greenfield LJ. Preclinical and clinical evaluation of a percutaneous stainless-steel Greenfield filter [abstr]. J Vasc Interv Radiol 1996; 7(suppl 1):195–196.
40. Xian ZY, Roy S, Hosaka J. Multiple emboli and filter function: An in vitro comparison of three vena cava filters. J Vasc Interv Radiol 1995; 6:887–893.
41. Awh MH, Taylor FC, Lu C-T. Spontaneous fracture of a Vena-Tech inferior vena cava filter. Am J Roentgenol 1991; 157:177–178.
42. Simon M, Rabkin DJ, Kleshinski S, et al. Comparative evaluation of clinically available inferior vena cava filters with an in vitro physiologic simulation of the vena cava. Radiology 1993; 189:769–774.
43. Qian Z, Yasui K, Nazarian G, et al. In vitro and in vivo experimental evaluation of a new vena caval filter. J Vasc Interv Radiol 1994; 5:513–518.
44. LaPlante JS, Contractor FM, Kiproff PM, et al. Migration of the Simon nitinol vena cava filter to the chest. Am J Roentgenol 1993; 160:385–386.
45. Roehm JOF, Thomas JW. The twist technique: A method to minimize wire prolapse during Bird's Nest filter placement. J Vasc Interv Radiol 1995; 6:455–459.
46. Vesley T, Darcy M, Picus D, et al. Technical problems associated with placement of the bird's nest inferior vena cava filter. Am J Roentgenol 1992; 158:875–880.
47. Aruny JE, Kandarpa K. Phlegmasia cerulea dolens, a complication after placement of a Bird's nest vena cava filter. Am J Roentgenol 1990; 154:1105–1106.
48. Nakawaga N, Cragg A, Smith TP, et al. A retrievable nitinol vena cava filter: Experimental and initial clinical results. J Vasc Interv Radiol 1994; 5:507–512.
49. Vorwerk D, Schmitz-Rode T, Schürmannk, et al. Use of a temporary caval filter to assist percutaneous iliocaval thrombectomy: Experimental results. J Vasc Interv Radiol 1995; 6:737–740.
50. Xian ZY, Roy S, Hosaka J, et al. In vitro evaluation of a new temporary venous filter: The Spring filter. Cardiovasc Intervent Radiol 1995; 18:315–320.
51. Kuszyk B, Venbrux AC, Samphilipo MA, et al. Subcutaneously tethered temporary filter: Pathological effects in swine. J Vasc Interv Radiol 1995; 6:895–902.
52. Stoneham GW, Burbridge BE, Millward SF. Temporary inferior vena cava filter: In vitro comparison with permanent IVC filters. J Vasc Interv Radiol 1995; 6:731–736.
53. Irie T, Yamauchi T, Makita K. Retrievable IVC filter: Preliminary in vitro and in vivo evaluation. J Vasc Interv Radiol 1995; 6:449–454.
54. Hicks ME, Malden ES, Vesley TM, et al. Prospective anatomic study of the inferior vena cava and renal veins: Comparison of selective renal venography and cavography and relevance in filter placement. J Vasc Interv Radiol 1995; 6:721–729.

Information for Authors

THE MOUNT SINAI JOURNAL OF MEDICINE, established in 1934 as *The Journal of the Mount Sinai Hospital,* is a peer-reviewed general medical journal. It is indexed or abstracted in *Index Medicus, Current Contents, Science Citation Index, Hospital Literature Index, International Nursing Index, Excerpta Medica, Chemical Abstracts, Biological Abstracts,* and major databases. Theoretical and basic science articles for the clinician, reviews, reports of clinical or research experience, articles on any subject affecting the practice of medicine, short notes in any specialty, book reviews, letters, and articles by medical students are welcome. The *Journal* is a participant in the Agreement on Uniform Requirements for Manuscripts Submitted to Biomedical Journals. The *Journal* aims at responding to authors of manuscripts within 8 weeks of submission.

Communications. Address correspondence on manuscripts to: Editor, *The Mount Sinai Journal of Medicine,* The Mount Sinai School of Medicine, Box 1094, One East 100th Street, New York, NY 10029-6574. Phone: (212) 241-6108; FAX: (212) 722-6386; e-mail: msjournal@msvax.mssm.edu.

Authorship Responsibility, Financial Disclosure, and Assignment of Copyright. Each author is required to sign (1) the statement on authorship responsibility; (2) the statement on financial disclosure; and (3) either the statement on copyright transfer or the statement indicating that the author is or was a US government federal employee when this work was conducted and prepared for publication. These statements are printed from time to time on a page in the *Journal* entitled "Required Form for Authors" and can be photocopied for use. Copies may also be obtained from the *Journal* office. Use additional copies of the form if there are more than six authors, and replace with the appropriate number of the author, e.g., 7th, 8th, etc. The signed original form(s) should be submitted with the manuscript.

Manuscript Preparation. Submit manuscripts in triplicate, including three sets of illustrations (see below). **Manuscripts must be typewritten or computer generated, double spaced throughout (including references, legends, and tables),** on one side of 8½ × 11 inch white bond paper with generous margins. All pages except the title page should be numbered consecutively. Begin a new page for the title page, abstract, initial page of text, references, each table, and legends. A floppy disk, if available, should accompany the manuscript. The label on the disk should specify the name of the first author, the title of the manuscript, how the disk has been formatted, and the word processing system used (DOS, Windows, or Windows 95). PC format is preferred but a high-density formatted Macdisk is acceptable. The text should be in ASCII, Microsoft Word or Wordperfect. Other than the title page, the abstract page and each page of text should be numbered sequentially, and include a heading (top right) which identifies the last name of the first author.

Drugs must be identified by their generic names. Do not use abbreviations on the title or abstract page. Define all abbreviations on first use in the body of the manuscript.

Any material taken directly from other sources must be accompanied by a written statement from both the author and publisher giving permission to the *Journal* for reproducing it. The authors must document in the manuscript that all clinical investigations described in the manuscript were conducted in accord with the guidelines of the Declaration of Helsinki and have been approved by the Institutional Review Board or its equivalent. The text must indicate that the informed consent of each subject was obtained. If results of experiments on animals are reported in the manuscript, the text must indicate that these studies were conducted in accord with the National Institutes of Health (US) guidelines for the care and use of laboratory animals and have been approved by the institutional office or its equivalent which supervises such work.

Title Page. The title page should be double spaced, and not numbered. It should indicate (1) title of the article, phrased as concisely as possible; (2) last and first name of each author, middle initials, if any, and not more than two academic degree(s); (3) institutional affiliation and title or rank of each author; (4) name and address of department(s) and institution(s) from which the work originated; (5) the name and address (and FAX number if available) of the author who will receive and respond to correspondence; (6) the name and full postal address of the author from whom reprint(s) are to be requested; (7) acknowledgments of sources of grant support; (8) name, place, and date of any conferences at which some or all of the information was presented; and, (9) a running head of fewer than 40 letter spaces. At the bottom of the title page, *list 3 to 10 key words or phrases* to facilitate indexing in accord with Medical Subject Headings from Index Medicus.

Abstract. All articles must include an abstract of not more than 250 words on the first numbered page of the manuscript. Abstracts should be double spaced and divided into four paragraphs:

Background: what problem(s) was(were) addressed;
Methods: how the study was performed;
Results: the salient results;
Conclusions.

Key Words. Three to 10 key words or phrases must be provided at the bottom of the title page to facilitate indexing. (see above)

Units of Measurement. SI units are preferred, but values may be expressed in conventional units and conversion factors indicated. The range of normal values for the reporting laboratory must be indicated. When conventional units are used in figures or tables, the appropriate conversion factor(s) must be provided in legends or footnotes.

Editing. All papers will be edited to enhance conciseness and clarity without altering meaning and to ensure conformity with journal style. Authors whose primary language is not English should obtain help from someone proficient in English spelling and grammar.

Illustrations. Photographs, line drawings, and graphs should enhance understanding of the text. Three copies of each must be submitted with the manuscript. Line drawings and graphs should be professionally prepared or computer generated and camera ready. Symbols, lettering, and numbers should be large enough to remain easily legible after the figure has been reduced to fit the width of one column. Photographs must be sharp, high contrast, black and white glossy prints not exceeding 5 inches in height × 7 inches in width. The authors must pay attention to the quality of print reproductions of x-ray plates or transparencies to assure that the special features are labeled and will remain easily visible when reduced to fit the width of one column. The name of the subject and hospital number(s) should not be reproduced in the print. Black and white copies of photomicrographs are preferred and must be cropped to a width of 3⅛ inches. The type of stain and power of magnification must be indicated in the relevant legend. Electron photomicrographs should have internal scale markers. On the back of each print and photograph affix a gum label on which the number of the figure has been written to correspond to its identification in the text, title of manuscript, first author's last name, and the word "top" and arrow to indicate the top edge. **Do not write this information directly on the print and do not mount the print on cardboard.** Illustrations which are to appear together should be mounted accordingly as plates and should be equally sized. Protect the prints by enclosing them in cardboard and separate envelopes. **Do not use paper clips.**

Every effort should be made to avoid using a photograph which includes the face or part of the face of a subject. Facial features should be marked to prevent identification of the subject unless the illustration carries details absolutely essential to the text. In any case, a partial or full facial photograph of a

subject will not be published unless written permission has been obtained from the subject and accompanies the photograph.

Legends. Legends should not duplicate the text. Each should be on a separate sheet of paper and double spaced. Number each legend to match the illustration it describes. Illustrations mounted as plates to appear together should be accompanied by a single legend identifying the separate elements by position ("upper left," "middle") or by letters. If such illustrations have been labeled with letters, the same letters are to be used in the legend. The legend must indicate the magnification and type of stain used for photomicrographs.

Tables. Each table should be typed on a separate page, double spaced. Number tables consecutively (Table 1, etc.) to correspond to their appearance in the text. Give each table a brief title that is self-explanatory without reference to the text and type the title on a separate line below its number. Identify all abbreviations and/or symbols used in the table in a footnote.

References. All references cited in the text are to be numbered sequentially. Enclose the citation number in parentheses and place it before any punctuation, e.g., Cytomegalovirus (1), steroids (2,3), and recreational drugs (4–6) have all been implicated (7). Do not use superscripts. The reference list must be limited to works cited in the text, and numbered in the order in which they are cited. Type, double spaced on separate pages following the arrangement and punctuation style in the examples below. Journal title abbreviations must conform to those used in *Index Medicus*.

1. Brown LK. Sleep apnea syndromes: overview and diagnostic approach. Mt Sinai J Med 1994; 61:99–112.

2. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. N Engl J Med 1991; 324:424–428.

3. Couser WG. Glomerular diseases. In: Stein J. ed. Internal Medicine 2nd ed. Boston: Little Brown: 1987: 834–861.

List the name of all authors when there are four or fewer. When there are more than four, list the names of the first three, then *et al.*

References to personal communications and unpublished data should not be included in the list of numbered references. Instead, such information should be placed in the text in parentheses. Permission must be obtained from the appropriate individual to so quote the material. This permission must accompany the manuscript at the time of its submission. When the manuscript is accepted for publication, the author(s) will be asked to confirm the accuracy of all references.

Proofs. Proofs are sent to the corresponding author from the printer. Stylistic changes should not be made. Because of increased printing charges, the cost of author's alterations beyond routine correction of typographical errors in text and data will be billed to the author. Proofs should be corrected and returned to and received at the editorial office within 48 hours by express delivery at the author's expense to the address listed above ("Communications").

Reprint Orders. A reprint order form is sent to the corresponding author with proofs; if reprints are desired, the form must be returned to the editorial office, preferably with the proofs, but not later than 3 days to avoid late charges. A check for the exact amount in US dollars should be made out to the order of The Mount Sinai Journal of Medicine. Please indicate on the check the last name of the first author, and the title of the article. If for one reason or another, the reprint order form is not received with the proofs, it may be obtained directly from the editorial office via telephone or FAX. Reprints are available in minimum quantities of 100, at a cost depending on the number of printed pages and whether covers are desired. This information is directly available from the editorial office.

Required Form for Authors:

AUTHORSHIP RESPONSIBILITY, FINANCIAL DISCLOSURE, AND ASSIGNMENT OF COPYRIGHT

Article Title (type):

Each author must read and sign this form dealing with (1) authorship responsibility, (2) financial disclosure and (3) copyright transfer. An author who was a US federal employee when this work was conducted and prepared for publication, must sign statement (4) instead of statement (3).

1. Authorship Responsibility

I certify that I have participated sufficiently in the conception and design of this work, and the analysis of the data (where applicable), as well as the writing of the manuscript, to take public responsibility for it. I believe that the manuscript represents valid work. I have reviewed the final version of the manuscript and approve it for submission for publication. Neither this manuscript nor one with substantially similar content under my (our) authorship has been published or is being considered for publication elsewhere, except as described in a separate attachment. Furthermore, I agree to produce the original recorded data on which the manuscript is based for examination by the Editor-in-Chief and/or his assignees should such records be requested.

2. Financial Disclosure

I certify that I have no affiliation with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials being discussed in the manuscript (eg, employment, consultancies, stock ownership, honoraria), except as disclosed in a separate attachment.

Author(s) Signature(s) **Date**

1st ______________________________

2nd ______________________________

3rd ______________________________

Author(s) Signature(s) **Date**

4th ______________________________

5th ______________________________

6th ______________________________

3. Copyright

In compliance with the Copyright Revision Act of 1976, the signature of each author on this form shall indicate and be evidence of agreement and mutual understanding with *The Mount Sinai Journal of Medicine*, in consideration of the further action of the *Journal* to review and edit the manuscript being submitted, to transfer, assign, and convey all copyright ownership, including any and all rights incidental thereto, exclusively to *The Mount Sinai Journal of Medicine* if this work is published in the *Journal*.

Author(s) Signature(s) **Date**

1st ______________________________

2nd ______________________________

3rd ______________________________

Author(s) Signature(s) **Date**

4th ______________________________

5th ______________________________

6th ______________________________

4. US Federal Employees

I was an employee of the US federal government when this work was conducted and prepared for publication; therefore, it is not protected by the Copyright Act and copyright ownership cannot be transferred.

Author(s) Signature(s) **Date**

1st ______________________________

2nd ______________________________

3rd ______________________________

Author(s) Signature(s) **Date**

4th ______________________________

5th ______________________________

6th ______________________________

Return the original signed form to The Editor, The Mount Sinai Journal of Medicine, Mount Sinai School of Medicine, Box 1094, One East 100th Street, New York, N.Y. 10029. Retain one copy of the form for your file. You may make as many photocopies as needed if your manuscript has more than six authors.

LEVY LIBRARY, MT. SINAI MEDICAL CENTER
3 4805 04114514

THE
MOUNT SINAI JOURNAL
OF MEDICINE

Forthcoming:

Grand Rounds

Targeted Gene Disruption with Antisense DNA: Research and Therapeutic Applications
Alan M. Gewirtz
University of Pennsylvania School of Medicine

The Supraventricular Tachycardias
M. Noelle S. Langan
The Mount Sinai Medical Center

Peptic Ulcer Disease: Paradigms Lost
Mark Babyatsky
The Mount Sinai Medical Center

Androgen Receptor Mutations in Prostate Cancer
Evelyn R. Barrack
The Johns Hopkins University

Theme Issues

Computers in the Doctor's Office
Editors: Ray Cornbill and Joseph Kannry

Laparoscopy
Editors: Barry Salky and Michael Edye

Cross-Cultural Issues in Psychopharmacology
Editors: John Herrera and William Lawson

Emergency Medicine
Editors: Andrew Jagoda and Lynne Richardson

Available now:

Maternal and Infant Health: New Perspectives
Kenneth V. Lieberman and
Charles J. Lockwood, editors
volume 61, no. 5, October 1994
88 pages **$15**

Update on Sleep Disorders
Lee K. Brown, editor
volume 61, no. 2, March 1994
96 pages **$15**

Environmental and Occupational Medicine in Clinical Practice
Richard P. Wedeen, editor
volume 62, no. 5, October 1995
96 pages **$15**

The Mount Sinai Hospital
1852

MOUNT SINAI SCHOOL OF MEDICINE
OF THE CITY UNIVERSITY OF NEW YORK

THE MOUNT SINAI JOURNAL OF MEDICINE

VOLUME 63 NUMBERS 5 & 6 OCTOBER / NOVEMBER 1996

OCT -7 1996

Issue 5

Cross-Cultural Issues in Psychopharmacology

Guest Editors: John Herrera and William Lawson

Part B: Alumni Theme Issue

Issue 6

General Articles

STATEN ISLAND UNIVERSITY HOSPITAL
Affiliated with The Mount Sinai Hospital and Mount Sinai School of Medicine

Veterans Affairs

THE MOUNT SINAI JOURNAL OF MEDICINE

The Mount Sinai Journal of Medicine is published by The Mount Sinai Medical Center of New York and has the following affiliates: Bronx Veterans Affairs Medical Center, New York; Elmhurst Hospital Center, Queens, NY; Queens Hospital Center, Queens, NY; and Staten Island University Hospital, Staten Island, NY.

Editor-in-Chief

Sherman Kupfer, M.D.

Editor Emeritus

Lester R. Tuchman, M.D.

Associate Editors

Harriet S. Gilbert, M.D.   Julius Wolf, M.D.

Administrative Assistant

Joseline Ilagan

Assistant Editors

David H. Bechhofer, Ph.D.
Constanin A. Bona, M.D., Ph.D.
Edward J. Bottone, Ph.D.
Lewis Burrows, M.D.
Christopher P. Cardozo, M.D.
Mark R. Chassin, M.D., MPH, MPP
Joseph S. Eisenman, Ph.D.
Richard A. Frieden, M.D.
Frederick Friedman, M.D.
Steven Fruchtman, M.D.
Paul L. Gilbert, M.D.
James H. Godbold, Ph.D.
Richard S. Haber, M.D.
Noam Harpaz, M.D.
Tomas Heimann, M.D.
Sheldon Jacobson, M.D., FACEP
Barry W. Jaffin, M.D.
Andrew S. Kaplan, D.D.S.
Samuel Kenan, M.D.
Suzanne Carter Kramer, M.Sc.
Mark G. Lebwohl, M.D.
Zvi Lefkowitz, M.D.
Kenneth V. Lieberman, M.D.
Lynda R. Mandell, M.D., Ph.D.
Steven Markowitz, M.D.
Bernard Mehl, D.P.S.
Myron Miller, M.D.
Edward Raab, M.D.
Elliot J. Rayfield, M.D.
Allan Reed, M.D.
Clive Rosendorff, M.D., Ph.D.
Fred Rosner, M.D.
Allan E. Rubenstein, M.D.
David B. Sachar, M.D.
Henry Sacks, M.D.
Ira Sanders, M.D.
Martin H. Savitz, M.D.
Clyde B. Schechter, M.D.
Stuart C. Sealfon, M.D.
Michael Serby, M.D.
Phyllis Shaw, Ph.D.
Alex Stagnaro-Green, M.D.
Barry D. Stimmel, M.D.
Nelson Stone, M.D.
Max Sung, M.D.
Rein Tideiksaar, Ph.D.
Richard P. Wedeen, M.D.

Editorial Board

Barry S. Coller, M.D.
Barry Freedman, M.B.A.
Richard Gorlin, M.D.
Nathan Kase, M.D.
Panayotis G. Katsoyannis, Ph.D.
Charles K. McSherry, M.D.
Jack G. Rabinowitz, M.D.
John W. Rowe, M.D.
Alan L. Schiller, M.D.
Alan L. Silver, M.D.
Alvin S. Teirstein, M.D.
Rosalyn S. Yalow, Ph.D.

The Mount Sinai Journal of Medicine (ISSN No. 0027-2507; USPS 284-860) is published 6 times a year in January, March, May, September, October, and November in one indexed volume by The Mount Sinai Medical Center, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Subscription price: individuals (U.S. and Canada), $75 per year; libraries and institutions (U.S. and Canada), $115 per year; individuals (outside U.S. and Canada), $100 per year; libraries and institutions (outside U.S. and Canada), $140 per year. Single copies $20. New subscriptions begin with the first issue of the calendar year. Send notice of change of address 60 days before its effective date. Periodicals postage paid at New York, NY and at additional mailing offices. POSTMASTER: Send address changes to The Mount Sinai Journal of Medicine, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Telephone number (212) 241-6108; Fax (212) 722-6386. Copyright 1996 The Mount Sinai Hospital.

DISCLAIMER: The statement and opinions expressed in the *Journal* by authors and contributors are solely their own and not those of The Mount Sinai Medical Center or The Mount Sinai Journal of Medicine. All advertisements are expected to conform to ethical standards, and their appearance does not imply endorsement or approval by The Mount Sinai Medical Center and/or The Mount Sinai Journal of Medicine of the products with respect to effectiveness, quality, or safety. The Mount Sinai Medical Center and The Mount Sinai Journal of Medicine do not hold themselves responsible for statements made by the authors or contributors, and for any injury to persons or property which may have resulted from ideas, products, and services referred to in articles or advertisements.

COPYRIGHT: Material printed in the Journal is protected by copyright. In accord with U.S. Copyright law, no part of the Journal may be reproduced, displayed, or transmitted in any form, electronic or mechanical, including photocopying or by any information storage or retrieval system, without prior written permission to do so. Permission should be requested from The Editor. The Mount Sinai Journal of Medicine, Mount Sinai School of Medicine, Box 1094, One East 100th Street, New York, NY 10029.

THE
MOUNT SINAI JOURNAL
OF MEDICINE

Volume 63
Numbers 5 & 6
October/November 1996

CONTENTS—Issue 5

Cross-cultural Issues in Psychopharmacology

Psychopharmacology in Cross-cultural Psychiatry *Keh-Ming Lin* 283

Ethnicity and Pharmacogenetics *Michael W. Smith and Ricardo P. Mendoza* 285

Clinical Importance of Genetic Polymorphism of Drug Oxidation *Timi Edeki* 291

The Art and Science of the Psychopharmacotherapy of African Americans *William B. Lawson* 301

Neuroleptic Dosing in Asian and Hispanic Outpatients with Schizophrenia *Sigfried Ruiz, Peter Chu, John Sramek, Ecterina Rotavu, and John Herrera* 306

Neuroleptic Dosing in Hispanic and Asian Inpatients with Schizophrenia *Yasmine Collazo, Raymond Tam, John Sramek, and John Herrera* 310

Ethnicity and Movement Disorders *John H. Eastham, Jonathan P. Lacrao, and Dilip V. Jeste* 314

Ethnicity and Antidepressant Response *John J. Sramek and Edmond H. Pi* 320

Gender Differences in Psychotropic Medications *Freda Lewis-Hall* 326

Ethnicity and Psychopharmacology in Latin America *Luis F. Ramirez* 330

National Institute of Mental Health: An Update on Women and Minorities in Research *Rick A. Martinez* 332

The *Journal* thanks Janssen Pharmaceutica Research Foundation for an educational grant in partial support of the cost of publication of the articles listed above.

Part B: Alumni Theme Issue

Sarcoidosis Mythology *Alvin S. Teirstein, Maria L. Padilla, Louis R. DePalo, and Gregory J. Schilero* 335

continued

THE
MOUNT SINAI JOURNAL
OF MEDICINE

Volume 63
Numbers 5 & 6
October/November 1996

CONTENTS *continued*

Fifty Years of Experience with Antithrombotic Therapy in Cardiac Disease: A 1996 Approach Based on Pathogenesis and Risk *Ronald S. Freudenberger and Valentin Fuster* 342

Congress and the International Medical Graduate: The Need for Equity *Barry D. Stimmel* 359

Isoniazid Hepatotoxicity after Orthotopic Liver Transplantation *Leona Kim Schluger, Patricia A. Sheiner, Mark Jonas, James V. Guarrera, Isabel M. Fiel, Burt Meyers, and Paul D. Berk* 364

The *Journal* thanks The Mount Sinai Alumni for an educational grant in partial support of the cost of publication of The Alumni Theme Issue, Parts A & B.

CONTENTS—Issue 6

Thomas C. Chalmers—In Memoriam *James F. Holland* 370

FROM BENCH TO BEDSIDE AND BACK

Perturbing Gene Expression with Oligodeoxynucleotides: Research and Potential Therapeutic Applications *Alan M. Gewirtz* 372

DEAN'S SEMINAR

G Proteins: Out of the Cytoskeletal Closet *Martin Rodbell* 381

GRAND ROUNDS

Peptic Ulcer Disease: Paradigms Lost *Andrea Sarner and Mark Babyatsky* 387

GENERAL ARTICLES

The Osteoclast and Osteoporosis *Steven L. Teitelbaum* 399

Androgen Receptor Mutations in Prostate Cancer *Evelyn R. Barrack* 403

Childhood Immunizations—1996 *David S. Hodes* 413

Commentary: Physician's Reactions to Patients: What Has Happened during the Past 10 Years *Richard Gorlin, James Strain, and Rosamond Rhodes* 420

Could Education Be Dangerous to Your Health? Pitfalls of Misapplied Statistics *Laszlo Sarkozi, Ellis Jacobs, Donald A. Smith, and Arthur Nizza* 425

DERMATOLOGY NOTES

Genital Herpes *Donald Rudikoff and Mark Lebwohl* 428

THE
MOUNT SINAI JOURNAL
OF MEDICINE

Volume 63
Numbers 5 & 6
October/November 1996

CONTENTS *continued*

BOOK REVIEWS

Effects of Atomic Radiation: A Half-Century of Studies from Hiroshima and Nagasaki *Barry S. Rosenstein* 429

The Cure of Childhood Leukemia: Into the Age of Miracles *Janet Cuttner* 430

Dermatologic Clinics *Sherman Kupfer* 430

Commentary: Introductory Remarks on the Code of Maimonides *Barry S. Coller; some notes by Fred Rosner* 432

List of Journal Reviewers in 1996 435

Author, Subject, Title Index to Volume 63 437

Psychopharmacology in Cross-cultural Psychiatry

Keh-Ming Lin, M.D., M.P.H.

Until the last few years, suggestions regarding possible cross-cultural/cross-ethnic differences in psychotropic responses were typically brushed aside as either unlikely or unimportant. Reflecting such a general attitude, very few designed research efforts had been made to examine issues in this area. Consequently, remarkable recent progress notwithstanding, much remains unknown regarding the influence of ethnic and cultural factors on the practice of modern psychopharmacology. This is ironic and regrettable. Without exception, all of our modern-day psychopharmacotherapeutic armamentaria are powerful substances with the potential of performing miraculous healing, yet are also potentially toxic and could be harmful when administered inappropriately. Since the 1950s, these drugs have been widely prescribed in this country for millions of patients with unique ethnic and cultural backgrounds. At the same time, the introduction of psychopharmacotherapy into non-Western countries has met almost no resistance, and an even larger number of psychiatric patients with extremely divergent genetic endowment, lifestyles, and cultural practices have been exposed to these modern-day wonder drugs which have helped to revolutionize the care of mental patients worldwide. The widespread use of these therapeutic agents attests to their efficacy and utility. At the same time, however, it is disconcerting to realize that, despite remarkable success in the practice of psychopharmacology across cultural and ethnic lines, so little is currently known regarding the potential contribution of cultural and ethnic factors to determining whether a particular patient would benefit from a particular treatment regimen, or if he or she is at a particular risk for certain potentially dangerous or devastating adverse effects attributable to the medication.

One can think of many reasons that contributed to such an unfortunate state of affairs. First of all, contrary to strong evidence indicating otherwise, a pervasive belief in psychiatry equates biology with universality and attributes cross-cultural/cross-ethnic diversity only to psychosocial factors (1, 2). This neglect of the remarkable biological diversity that typically exists in any living organism, including the human species, both within and between ethnic groups, has been at least partially responsible for the slow progress of cross-cultural research on biologically related factors. Because of historical antecedents (3), researchers and clinicians are often uncomfortable with the idea that people with divergent ethnic and ancestral backgrounds might differ significantly in their biological endowments, including pharmacological responses. This "color-blind" approach has resulted too often in the neglect of issues or concerns that may be particularly important for ethnic minority populations; this approach has also resulted in the imposition of guidelines and norms based on experiences with white males on the great majority of patients with divergent backgrounds. Perhaps, partially as a reflection of the biases stated above, there has been a deep-rooted dichotomy between clini-

---

This work was supported in part by the National Institute for Mental Health Research Center on the Psychobiology of Ethnicity MH47193 Grant.

From the Department of Psychiatry, University of California, Los Angeles, and Research Center on the Psychobiology of Ethnicity, Harbor-UCLA Medical Center, Torrance, California. Address reprint requests to Keh-Ming Lin, M.D., M.P.H., Director, Research Center on the Psychobiology of Ethnicity, Harbor-UCLA Medical Center, 1124 W. Carson St., Torrance, CA 90502.

cians and researchers engaged in working with cross-cultural and ethnic minority populations on the one hand, and biological researchers on the other, with the former primarily interested in psychosocial and health service issues, and the latter paying little attention to ethnicity and culture. Despite consensus regarding the importance of "biopsychosocial" integration in psychiatric practice and research, compartmentalization is pervasive, which results in a regrettable loss of opportunity for cross-fertilization and synergy. This is particularly problematic for research involving ethnic minority patients.

Similarly deep-rooted dichotomy between basic and clinical researchers exists, limiting interdisciplinary collaboration and exchange of information. For example, although progress in the field of pharmacogenetics has been truly remarkable in recent years, and many of the exciting new findings could potentially explain some of the clinical observations of ethnic differences in psychotropic responses, very few of these have been available in psychiatric journals. Although several recently developed pharmacogenetic probes, which use both biochemical and molecular biology methods, show promise for predicting interindividual and cross-ethnic differences in drug responses, they have not been tried in the clinical settings.

Last, the equally sizable interindividual variation in pharmacological responses could also exert a masking effect on the existence of cross-ethnic differences, which typically superimposes on the former. The graph (Fig. 1) shows: 1) substantial interindividual variability within each of the ethnic groups; 2) dramatic differences in the pharmacokinetics of haloperidol between the two ethnic groups; and 3) overlap of the pharmacokinetics between the two groups. Ethnic and cultural comparisons frequently are cast in simplistic terms without simultaneously paying attention to the reality of the substantial variability that also exists within each of the groups. This simplistic approach to ethnic comparisons poses two opposing but related problems. On the one hand, reports of significant cross-ethnic/cross-cultural differences in psychotropic responses risk being interpreted stereotypically, leading to all patients from one of the groups *always* being treated with lower doses of a particular medicine, demonstrating a reluctance to tailor the treatment regimen individually. On the other hand, the existence of substantial overlap in drug responses between the comparison groups may lead one to lose sight of the fact that ethnicity does determine for most patients how they might respond to a particular medication.

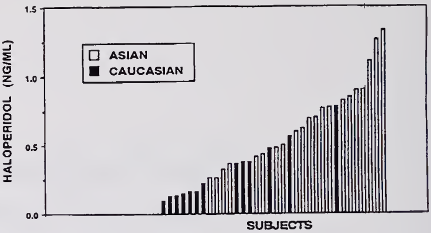


Fig. 1. Variability of haloperidol concentrations in plasma of normal volunteers after the administration of haloperidol (1.0 mg p.o.).

Despite these potential pitfalls and obstacles, remarkable progress has been made in recent years. This is clearly demonstrated by the excellent quality of the contributions included in this special issue. Although this is still a young field, and much remains to be clarified, the information presented and reviewed in the following papers is intriguing, thought provoking, and clinically relevant. It is clear that, with the collective talent represented here and elsewhere, research in this area will grow exponentially, ensuring that the practice of psychopharmacotherapy will be increasingly rational, scientifically based, and culturally sensitive.

References

1. Kleinman A. Rethinking psychiatry. New York: Free Press, 1988.
2. Lin KM, Poland RE, Nakasiki G. Psychopharmacology and psychobiology of ethnicity. Washington, DC: American Psychiatric Press, 1993.
3. Lawson W. Racial and ethnic factors in psychiatric research. Hosp Commun Psychiatry 1986; 37:50–54.

Ethnicity and Pharmacogenetics

Michael W. Smith, M.D. and Ricardo P. Mendoza, M.D.

Abstract

Ethnic differences in pharmacological response have been recognized for a long time. New research methodologies in the field of pharmacogenetics have begun to provide us with important insights concerning the biological mechanisms that underlie this differential response. The emerging data have highlighted the central role of genetic factors in the metabolism, and perhaps plasma protein binding, of many of the psychotropic agents used today. Ethnic differences in these genetic factors may explain in large measure why patients from ethnic minorities respond differently than Caucasians when given psychotropic medications. However, it is clear that, as for psychiatry in general, the role of the environment in these genetic factors must be understood when considering psychopharmacological response in particular. The utility and possible application of these research methodologies in the clinical setting is still undetermined; these data must be pursued. The information concerning the relative efficiency of the drug-metabolizing enzymes which is made available through genotyping and phenotyping methodologies could be used by clinicians who provide psychopharmacotherapeutic services to patients from diverse ethnic backgrounds.

RESEARCH IN THE FIELD of pharmacogenetics has recently begun to answer many of the questions that have troubled clinicians who are involved in the psychopharmacotherapeutic management of patients from ethnic minorities. Since the advent of biological treatment strategies in psychiatry in the 1960s, numerous clinical reports and anecdotal observations have surfaced which suggest that patients from ethnic minorities respond differently than their Caucasian counterparts when given psychotropic medications. Pharmacokinetic research has produced compelling evidence that validates these early reports, but the biological mechanisms that underlie this differential response have remained elusive. In the past decade, pharmacogenetic research (the study of genetically linked idiosyncratic data responses) has highlighted the central role of genetic factors in the metabolism of many of the psychotropic agents used today (1–4). Ethnic differences in the genetic structure of the drug-metabolizing enzymes may explain, in large measure, the ethnic differences that have been reported in pharmacological response.

This article closely examines the relationship of ethnicity and pharmacogenetics, reviewing the data regarding variations in genetic structure of the drug-metabolizing enzymes and the resultant differences in the functional activity of these enzymes. It also summarizes the data concerning the various environmental factors that have been shown to influence the activity and efficiency of the drug-metabolizing enzymes. Because of the ongoing demographic shifts in the population of the United States, U.S. psychiatrists probably will be routinely involved in the pharmacological management of patients from ethnic minorities. Therefore, the clinical implications of these find-

---

This work was supported, in part, by the Research Center on the Psychobiology of Ethnicity Grant MH47193.

From the Department of Psychiatry, University of California, Los Angeles, Research Center on the Psychobiology of Ethnicity, Harbor-UCLA Research and Education Institute, Torrance, CA. Address reprint requests to Michael W. Smith, M.D., Assistant Clinical Professor of Psychiatry, University of California, Los Angeles, Research Center on the Psychobiology of Ethnicity, Harbor-UCLA Research and Education Institute, Bldg. B-4 South, 1124 W. Carson St., Torrance, CA 90502.

ings for practicing psychiatrists will also be discussed.

Pharmacogenetic Research and the Cytochrome P-450 Enzyme System

The cytochrome P-450 (CYP) enzyme system is the main pathway of drug metabolism in man. In numerous *in vitro* and *in vivo* studies (largely with healthy volunteers), these enzymes have been demonstrated to be responsible for the metabolism of the majority of psychotropics currently in use, as well as those that will soon be on the market (5–10). Because of its importance in drug metabolism, the CYP system has recently been the focus of researchers in the field of pharmacogenetics to answer questions concerning ethnicity and psychopharmacological response. Ethnic differences in drug metabolism appear to be related to polymorphic (different forms of the same enzyme) variations in the CYP enzymes. This polymorphic variability may be the result of evolutionary forces, primarily molecular drive, acting on the CYP system. Early terra forms encountered many potentially harmful substances in their diet (e.g., toxins in edible plants) and environment. Throughout evolution CYP enzymes have exhibited high mutagenicity; this plasticity proved protective because species survival depend on the efficiency of these metabolic enzymes (2). A major clinical consequence of CYP evolution, however, has been CYP enzyme polymorphism, and many idiosyncratic drug reactions have been linked to these polymorphic enzymes (11–15).

The two most extensively studied CYP enzymes are debrisoquin hydroxylase H (CYP2D6) and mephenytoin hydroxylase (CYP2C19). CYP2D6 and CYP2C19 are responsible for the metabolism of many commonly used medications such as cardiovascular agents, analgesics, and psychotropic medications (3–4, 16, 17). Individuals possessing polymorphic forms of these enzymes can be classified as either extensive metabolizers (EMs) or poor metabolizers (PMs) (1, 18, 19). PMs appear to lack a functional form of the enzyme which often results in much slower rates of drug metabolism and potentially toxic blood levels after the administration of standard doses. Reports also have demonstrated that the functional status of some of these enzymes may be associated with risks for conditions such as Parkinsonism (20) and several different malignancies (21–23). Substantial cross-ethnic differences exist in the frequency of CYP2D6 and CYP2C19 PMs. The frequency of PMs of debrisoquin varies from less than 3% in the Cuna Amerindians, Middle Easterners, Mexican Americans, and Asians (24–30), to 3–10% in Caucasians in Europe and North America (31). Among various groups of Black Africans, there is a wide range of frequencies, with 0–8% of Saharan Africans, 4% of Venda in South Africa, 1.9% of African Americans, and 19% of Sans Bushmen being classified as PMs (31, 32).

Recent genotyping studies (gene sequencing of the actual drug-metabolizing enzyme) have revealed genetic differences in the sequencing of the CYP enzymes among ethnically diverse individuals (1, 32). In addition, a specific point mutation has been identified in the majority of Caucasian PMs (approximately 7% of the population) of debrisoquin that gives rise to a completely inactive and nonfunctional enzyme. However, extensive metabolizers and poor metabolizers are relative terms and are not necessarily quantitatively or qualitatively comparable across ethnic groups (33). For example, although the frequency distribution of Asian PMs is only less than 1%, another 34% of Asians exhibit a metabolic capacity that is significantly lower than Caucasian EMs and would appear to represent a group of "slow" metabolizers (SMs) among this population. Genetic sequencing of these SMs reveals an abnormality in a large gene fragment of the enzyme (34). Caucasian PMs also have an abnormality in this same fragment; however, the variation in the group represents a point mutation as compared with a gene insertion among the Asian subjects (35, 36). This difference between point mutation and gene insertion produces isoenzymes that functionally behave entirely differently, namely PMs in Caucasians and SMs in Asians. In a recent study with African Americans, a high prevalence (33%) of the same genetic difference related to "slow" metabolism in Asian subjects was discovered (37). The frequency of PMs of mephenytoin hydroxylase also varies substantially across ethnic groups. Although PMs of this enzyme are relatively rare among Caucasians, they are quite prevalent in Asian populations, with approximately 20% of Japanese and Chinese being classified as PMs (31). A recent study on African Americans also reported a rate significantly higher than in Caucasians (33).

Pharmacogenetic Research and Plasma Protein Binding

Of the various proteins and cells in the blood that provide binding sites for drug compounds, only one category of plasma proteins, α_1-acid glycoproteins, have been shown to exhibit ethnic differences (38, 39). Ethnic variation is found in the absolute concentration of α_1-acid glycoprotein, as well as the distribution of its two variants (40–42). One of the variants appears to bind more readily to several psychoactive agents (nortriptyline, amitriptyline, and methadone), and varies in frequency from 15 to 27% in Asians to 34 to 67% in Caucasians and African Americans (41). The majority of psychotropic agents act on the central nervous system and must cross the blood-brain barrier to attach to central nervous system receptors. Because it is only the nonbound fraction of drugs that can cross the blood-brain barrier, variations in the concentration of these drug-binding proteins in the plasma can significantly influence the observed clinical effects, including side effects (43–46). To date, very few cross-ethnic studies of plasma protein binding have been conducted (47). They have yielded conflicting findings, and more systematic research in the area is needed.

The Clinical Application of Phenotype and Genotype

Phenotyping and genotyping methodologies currently are used primarily in research protocols. If practicing psychiatrists could access similar detailed information regarding the metabolic efficiency of their patients, the pharmacotherapeutic decision-making process might be greatly improved. The clinical application of phenotyping and genotyping methodologies, however, must be based on an understanding of their respective advantages and disadvantages. For example, genotypes do not appear to be influenced by environmental factors and are stable over the lifetime of an individual (48–50); this suggests that they could be used for "fingerprinting" purposes. Unfortunately, a molecular biology laboratory is needed to perform the test. Although phenotyping is performed much more easily, the interpretation of results may be difficult because the functional expression of drug-metabolizing enzymes can be altered by a variety of environmental influences. Despite recent evidence indicating that these influences may be far more extensive than previously believed (51–53), they have rarely been the focus of rigorous scientific investigation. The clarification of these issues is of considerable clinical importance, and research studies using methodologies that simultaneously assess genotypic and phenotypic expression, under a variety of clinical conditions, must be undertaken. The information obtained from these studies will provide a better understanding of the nature of the relationship between phenotype and genotype and will be useful in providing guidelines regarding their clinical utility.

Phenotypic Expression and Dietary and Environmental Influences

The activity of cytochrome P-450 enzymes, as well as other drug-metabolizing enzymes, can be influenced by exposure to different diets, environmental toxins, gender, age, disease processes, the medications used to treat these processes, and, finally, a wide variety of biological substrates (7, 54–62). These influences might also contribute to the observed ethnic differences in pharmacological response. The relationship between diet and enzymatic phenotypic expression has been the subject of several scientific investigations. Branch et al. (58) compared the rate of the metabolism of antipyrine and found a significantly longer half-life of the drug among Sudanese living in their home villages compared to Sudanese residing in Britain or to White British subjects. The latter two groups metabolized the antipyrine at similar rates, suggesting that environmental factors such as diet were responsible for the pharmacokinetic differences. Similar findings underscoring the importance of diet in phenotypic expression in Asian Indian subjects have been reported by other researchers (63–65).

A wide variety of biological compounds have been shown to both inhibit, as well as induce, the activity of many of the CYP isozymes (33, 66). These biological compounds include not only extremely potent research drugs, but also commonly used pharmaceutical agents (e.g., cimetidine and carbamazepine), as well as herbal medicines, chemical toxins, steroid and sex hormones, alcohol, caffeine, tea, grapefruit juice, and certain constituents of tobacco (51–67). Contrary to the beliefs of most physicians, many herbal drugs are pharmacologically active, and capable of significant interactions with prescribed drugs, both pharmacokinetically and pharmacodynamically (51, 53, 68–71). Several Chinese herbs, including muscone, ginseng, and glycyrrhiza, have been found to have potent stimulating effects on the cytochrome XP-450 enzymes. In contrast, oleanolic acid contained in *Sertia mileensis* and *Ligustrum lucidium* it substantially inhibits the activities of these enzymes (52). Clinicians who provide psychiatric services to patients from ethnic minorities should recognize that traditional herbal medicines continue to be used extensively by these patients (72–74), often side-by-side with modern Western pharmaceutical agents.

Psychotropic medications themselves exert influences on the drug-metabolizing enzymes. For example, the addition of the popular selective serotonin reuptake inhibitor antidepressant, fluoxetine, to treatment with the tricyclic antidepressant desipramine, has been reported to produce toxic levels of the tricyclic because of competitive inhibition of CYP2D6 (75). Competitive inhibition of CYP2D6 may also be responsible for the 46% rate of PMs observed among neuroleptic-treated patients as compared with the 3 to 9% rate of PMs in drug-free healthy subjects (76). Presently, it is not clear which neuroleptics predominantly inhibit the CYP2D6 enzyme and whether, and to what extent, there is a dose-effect relationship. The influence of concurrent psychotropic medications on phenotypic expression is of popular concern in light of the widespread use of "polypharmacy" in the United States and warrants further investigation.

Summary

Minority populations in the United States continue to increase and the growing numbers guarantee that American psychiatrists will be routinely involved in their care. The literature contains ample evidence attesting to the fact that ethnicity represents an important factor in determining the response of patients to a wide variety of medications and xenobiotics. Technological advancements in medicine have given rise to new research methodologies in the field of pharmacogenetics. These are being used to clarify the underpinnings involved in the differential response that patients from ethnic minorities display in response to the administration of psychotropic medications when compared with Caucasian counterparts. Pharmacogenetic research data indicate that ethnic differences in the genetic makeup of the drug-metabolizing enzymes, the cytochrome P450 system, and perhaps the plasma proteins that bind psychotropic agents are central in understanding this differential response. Because the genetic structure of these enzymes dictates their functional status, these factors have major clinical significance. However, the environment has been shown to exert powerful influences on the biotransformation of psychotropic medications and its role in psychopharmacological response must be appreciated. Currently, phenotyping and genotyping methodologies are largely research tools. In the future, studies should be undertaken to examine more closely the relationship between genotyping and phenotyping methodologies under a variety of clinical conditions. Establishing the appropriate means to apply these methodologies clinically could help guide the psychopharmacological decisions of practicing psychiatrists providing care to patients from ethnic minorities.

References

1. Gonzalez F. The molecular biology of cytochrome P450s. Pharmacol Rev 1989; 40:243–288.
2. Gonzalez F, Nebert D. Evolution of the P450 gene superfamily: animal-plant "warfare" molecular drive and human genetic differences in drug oxidation. Trends Genet 1990; 6:182–186.
3. Kalow W. Interethnic variation of drug metabolism. Trends Pharmacol Sci 1991; 12:102–107.
4. Meyer U, Zanger U, Grant D, Blum M. Genetic polymorphisms of drug metabolism. Adv Drug Res 1990; 19:12–17.
5. Bertilsson L, Eichelbaum M, Mellstrom B. Nortriptyline and antipyrine clearance in relation to debrisoquine hydroxylation in man. Life Sci 1980; 27:1673–1677.
6. Bertilsson L, Aberg-Wistedt A. The debrisoquine hydroxylation test predicts steady-state plasma levels of desipramine. Br J Clin Pharmacol 1983; 15:388–390.
7. Clark W, Brater D, Johnson A. Goth's medical pharmacology. 12th ed. St. Louis: The C.V. Mosby Company, 1988.
8. Kalow W. Race and therapeutic drug response. N Engl J Med 1989; 320:588–589.
9. Shen W, Lin K. Cytochrome P-450 monooxygenases and interactions of psychotropic drugs. Int J Psychiatry Med 1990; 21:21–30.
10. Yue QY, Hasselstrom J, Svensson JO, Sawe J. Pharmacokinetics of codeine and its metabolites in Caucasian healthy volunteers: comparisons between extensive and poor hydroxylators of debrisoquine. Br J Clin Pharmacol 1991; 31:635–642.
11. Bertilsson L, Mellstorm B, Sjoqvist F, Martensson B, Asberg M. Slow hydroxylation of nortriptyline and concomitant poor debrisoquine hydroxylation: clinical implications, Lancet 1981; 111:560–561.
12. Shah RR, Oates NS, Idle JR, et al. Impaired oxidation of debrisoquine in patients with perhexiline neuropathy. Br Med J 1982; 284:295–299.
13. Sjoqvist F, Bertilsson L. Slow hydroxylation of tricyclic antidepressants: relationship to polymorphic drug oxidation. In: Kalow W, Goedde HW, Agarwal DP, editors. Ethnic differences in reaction to drugs and xenobiotics. New York: NY: Alan R. Liss Inc., 1986: 169–188.
14. Dahl-Puustinen ML, Liden A, Alm C. Disposition of perphenazine is related to polymorphic debrisoquine hydroxylation in human beings. Clin Pharmacol Ther 1989; 6:78–81.
15. Dahl ML, Bertilsson L. Genetically variable metabolism of antidepressants and neuroleptic drugs in man, Pharmacogenetics 1993; 3:61–70.
16. Wood A, Zhou H. Ethnic differences in drug disposition and responsiveness. Clin Pharmacokinet 1991; 20:1–24.
17. Zhou H, Kashakji R, Siolberstein D, Wilkinson G, Wood A. Altered sensitivity to and clearance of propranolol in men of Chinese decent as compared with American white. N Engl J Med 1989; 320:565–570.
18. Kupfer A, Preisig R. Pharmacogenetics of mephenytoin: a

new drug hydroxylation polymorphism in man. Eur J Clin Pharmacol 1984; 26:753–759.

19. Wilkinson GR, Guengerich FP, Branch RA. Genetic polymorphism of S-mephenytoin hydroxylation. Clin Pharmacol Ther 1989; 43:53–76.
20. Smith CAD, Gough AC, Leigh PH, et al. Debrisoquine hydroxylase gene polymorphism and susceptibility to Parkinson's disease. Lancet 1992; 339:1375–1377.
21. Fleming CA, Kaisary A, Wilkinson GR, et al. The ability to 4-hydroxylate debrisoquine is related to recurrence of bladder cancer. Pharmacogenetics 1992; 2:128–134.
22. Kaisary A, Smith P, Jacqz E, et al. Genetic predisposition to bladder cancer: ability to hydroxylate debrisoquine and mephenytoin as risk factors. Cancer Res 1987; 47: 5488–5493.
23. Hirvonen A, Husgafvel-Pursiainen, Antila S, et al. PCR-based CYP2D6 genotyping for Finnish lung cancer patients. Pharmacogenetics 1993; 3:19–27.
24. Jorge LF, Arias TD, Inaba T, Jackson PR. Unimodal distribution of the metabolic ratio for debrisoquine in Cuna Amerindians of Panama. Br J Clin Pharmacol 1990; 30:281–285.
25. Mahgoub A, Idle JR, Smith RL. A population of familial study of the defective alicyclic hydroxylation of debrisoquine among Egyptians. Xenobiotica 1979; 9:51–56.
26. Islam SI, Idle JR, Smith RL. The polymorphic 4-hydroxylation of debrisoquine in a Saudi-Arab population. Xenobiotica 1980; 10:819–825.
27. Lou Y, Ying L, Bertilsson L, Sjoqvist. Low frequency of slow debrisoquine hydroxylation in a native Chinese population. Clin Pharmacol Ther 1987; 15:443–450.
28. Lee EJ, Nam YP, Hee GN. Oxidation phenotyping in Chinese and Malay populations. Clin Exp Pharmacol Physiol 1988; 15:889–891.
29. Xu X, Jian GW. Debrisoquine hydroxylation and sulfamethazine acetylation in a Chinese population. Chung-Kuo Yao Li Hsueh Pao (Acta Pharmacol Sin) 1990; 11:387–388.
30. Du Y, Lou Y. Polymorphism of debrisoquine 4-hydroxylation and family studies of poor metabolizers in Chinese population. Chung-Kuo Yao Li Hsueh Pao (Acta Pharmacol Sin) 1990; 11:7–10.
31. Silver B, Poland R, Lin K. Ethnicity and the pharmacology of tricyclic antidepressants. In: Lin PK, Poland R, Nakasaki G, editors. Psychopharmacology and psychobiology of ethnicity. Washington DC: American Psychiatric Press, 1993.
32. Masimirembwa CM, Johansson I, Hasler JAL, Magnus. Genetic polymorphism of cytochrome P450 CYP2D6 in Zimbabwean population. Pharmacogenetics 1993; 3: 275–280.
33. Meyer U. Molecular genetics and the future of pharmacogenetics. In: Kalow W, editor. Pharmacogenetics of drug metabolism. New York: Pergamon Press, 1992.
34. Wang SL, Huang JD, Lai MD, Liu BH, Lai ML. Molecular basis of genetic variation in debrisoquine hydroxylation in Chinese subjects: polymorphism in RFLP and DNA sequence of CYP2D6. Clin Pharmacol Ther 1993; 53: 410–418.
35. Kagimoto M, Heim M, Kagimoto K, Zeugin T, Meyer UA. Multiple mutations of the human cytochrome P4502D6 gene (CYP2D6) in poor metabolizers of debrisoquine. J Biol Chem 1990; 265:17206–17214.
36. Ingelman-Sundberg M, Johansson I, Persson I, et al. Genetic polymorphism of cytochromes P450: interethnic differences and relationship to incidence of lung cancer. Pharmacogenetics 1992; 2:264–271.
37. Relling MV, Cherrie J, Schell MJ, et al. Lower prevalence of the debrisoquine oxidative poor metabolizer phenotype in American blacks versus white subjects. Clin Pharmacol Ther 1991; 50:308–313.
38. Baumann P, Eap C. Alpha-acid glycoprotein genetics. Biochemistry, physiological functions and pharmacology. New York: Alan R. Liss, 1988.
39. Kremer J, Wilting J, Janssen L. Drug binding to human alpha-1-acid glycoprotein in health and disease. Pharmacol Rev 1988; 40:1–45.
40. Fukuma Y, Kashimimura S, Umetsu K, et al. Genetic variation of alpha-2-HS-glycoprotein in the Kyushu district of Japan: description of three new rare variants. Hum Hered 1990; 40:49–51.
41. Juneja R, Weitkamp L, Straitil A. Further studies of the plasma, $\alpha\beta$-glycoprotein polymorphism: two new alleles and allele frequencies in Caucasians and in American Blacks. Hum Hered 1988; 38:267–272.
42. Umetsu K, Yuasa I, Nishimura H. Genetic polymorphisms of orosomucoid and alpha-2-HS-glycoprotein in a Philippine population. Hum Hered 1988; 38:287–290.
43. Routledge P. The plasma protein binding of basic drugs. Br J Clin Pharmacol 1986; 22:499–506.
44. Levy R, Moreland T. Rationale for monitoring free drug levels. Clin Pharmacokinet 1984; S1:1–9.
45. Baumann P, Eap C. Plasma monitoring of antidepressants: clinical relevance of the pharmacogenetics of metabolism and of acid glycoprotein binding. Biol Psychiatry 1991; 29:75–95.
46. Crabtree B, Jann M, Pits W. Alpha acid glycoprotein levels in patients with schizophrenia: effect of treatment with haloperidol. Biol Psychiatry 1991; 29:43A–48A.
47. Mendoza R, Smith M, Poland R, et al. Ethnic psychopharmacology the Hispanic and Native American perspective. Psychopharmacol Bull 1991; 27:449–461.
48. Bertilsson L, Dahl M, Edqqvist B, et al. Genetic regulation of the disposition of psychotropic drugs.In: Meltzer HY, Nerozzi D, editors. Current practices and future developments in the pharmacotherapy of mental disorders. Amsterdam: Elsevier Science Publishers, 1991: 73–80.
49. Bronsen K, Sindrup SH, Skjelbo E, et al. Role of genetic polymorphism in psychopharmacology, an update. Psychopharmacol Ser 1993; 10:199–211.
50. Lin KM, Poland RE. Ethnicity, culture, and psychopharmacology. In: Bloom FE, Kupfer DI, editors. Psychopharmacology: the fourth generation of progress. New York: Raven Press, 1995.
51. Jones B, Runikis A. Interaction of ginseng with phenelzine. J Clin Psychopharmacol 1987; 201–202.
52. Liu GT. Effects of some compounds isolated from Chinese medicinal herbs of hepatic microsomal cytochrome P-450 and their potential biological consequences. Drug Metab Rev 1991; 23:439–465.
53. Shader R, Greenblatt D. Bees, ginseng and MAOIs revisited (editorial). J Clin Psychopharmacol 1988; 8:325.
54. Anderson K, Conney A, Kappas A. Nutritional influences on chemical biotransformations in humans. Nutr Rev 1982; 40:161–169.
55. Anderson KE. Influences of diet and nutrition on clinical pharmacokinetics. Clin Pharmacokinet 1988; 14:325–346.
56. Anderson KE, McCleery RB, Vesell ES, Vickers FF, Kappas A. Diet and cimetidine induce comparable changes in theophylline metabolism in normal subjects. Hepatology 1991; 13:941–946.

57. Anderson KE, Kappas A. Dietary regulation of cytochrome P450. Annu Rev Nutr 1991; 11;141–167.
58. Branch R, Salih S, Homeida M. Racial differences in drug metabolizing ability: a study with antipyrine in the Sudan. Clin Pharmacol Ther 1978; 24:283–286.
59. Brosen K, Skjelbo E. Fluoxetine and norfluoxetine are potent inhibitors of P4011D6–the source of the sparteine/debrisoquine oxidation polymorphism (letter). Br J Clin Pharmacol 1991; 32:136–137.
60. Llerena A, Herraiz AG, Cobaleda J, et al. Debrisoquine and mephenytoin hydroxylation phenotypes and CYP2D6 genotype in patients treated with neuroleptic and antidepressant agents. Clin Pharmacol Ther 1993; 54:606–611.
61. Sigusch H, Hippius M, Henschel L, et al. Influence of grapefruit juice on the pharmacokinetics of a slow release nifedipine formulation. Pharmazie 1994; 49:522–524.
62. Smith M, Lin K, Mendoza R. "Non-Biological" issues affecting psychopharmacotherapy: cultural considerations. In Lin K, Poland R, Nakasaki G, editors, Psychopharmacology and psychobiology of ethnicity. Washington: American Psychiatric Press, 1993.
63. Fraser H, Mucklow J, Bulpitt C, et al. Environmental factors affecting antipyrine metabolism in London factor and office workers. Br J Clin Pharmacol 1979; 7: 237–243.
64. Dollery C, Faser H, Mucklow J. Contribution of environmental factors to variability in human drug metabolism. Drug Metab Rev 1979; 9:207–220.
65. Desai NK, Sheth UK, Mucklow JC. Antipyrine clearance in Indian villagers. Br J Clin Pharmacol 1980; 9:387–394.
66. Okey A. Enzyme induction in the cytochrome P-450 system. New York: Pergamon Press, 1992.
67. Murray M, Ready G. Selectivity in the inhibition of mammalian cytochromes P-450 by chemical agents. Pharmacol Rev 1990; 42:2–101.
68. Yamahara J, Kobayashi M, Matsuda H, et al. Anticholinergic action of Swertia Japonica and an active constituent. J Ethnopharmacol 1991; 33:31–35.
69. Egashira T, Sudo S, Murayama F, et al. Effects of kamikihi-to, a Chinese traditional medicine, on various cholinergic biochemical markers in the brain of aged rats. Nippon Yakurigaku Zasshi (Folia Pharmacol Jpn) 1991; 98:273–281.
70. Carbajal D, Casaco A, Arruzazabala L, et al. Pharmacological screening of plant decoctions commonly used in Cuban folk medicine. J Ethnopharmacol 1991; 33:21–24.
71. Lewis WH, Kennelly EJ, Bass GN, et al. Ritualistic use of the holly ilex guayusa by Amazonian Jivaro Indians. J Ethnopharmacol 1991; 33;25–30.
72. Kleinman A. Patients and healers in the context of culture. Berkeley: University of California, 1980.
73. Vermeer DE, Farrell RE. Nigerian geophagical clay: a traditional antidiarrheal pharmaceutical. Science 1985; 227:634–636.
74. Westermeyer J. Psychiatric care of migrants: a clinical guide. Washington DC: American Psychiatric Press, 1989.
75. Brosen K, Hansen JG, Nielsen KK, et al. Inhibition by paroxetine of desipramine metabolism in extensive but not in poor metabolizers of sparteine. Eur J Clin Pharmacol 1993; 344:349–355.

Clinical Importance of Genetic Polymorphism of Drug Oxidation

Timi Edeki, M.D., Ph.D.

Abstract

Certain individuals have a metabolic deficiency in the metabolism of debrisoquin, sparteine, dextromethorphan, and more than 30 other clinically important drugs. Examples of such drugs include tricyclic antidepressants, neuroleptics, selective serotonin reuptake inhibitors, β-adrenoceptor blockers, and antiarrhythmics. CYP2D6, the enzyme responsible for the metabolism of these drugs, is polymorphically distributed in different populations. Studies in different ethnic groups in particular demonstrate significant variation. CYP2D6 deficiency has important therapeutic consequences, such as increased side effects when medications that are substrates of this enzyme are prescribed for such individuals. To optimize drug therapy, physicians should therefore determine the metabolic capacity of their patients.

In the past three decades several genetic defects in the way the body handles drugs have been discovered. These defects which may be responsible for unusual drug reactions, can be divided into two types: 1) defects caused by altered pharmacokinetic handling of the drug, i.e., the effect of the body on the drug and 2) defects caused by altered tissue responsiveness; these are pharmacodynamic in origin, i.e., the effect of the drug on the body. Acetylation and oxidation are examples of two areas of drug metabolism in which genetic polymorphism has been studied extensively. One of the earliest references to oxidative drug polymorphism was by Kutt (2), who described deficient *p*-hydroxylation of phenytoin. Interethnic variation in the metabolism of phenytoin has been reviewed recently (3). Shahidi (4) reported a pedigree of two sisters who exhibited defective *O*-deethylation of phenacetin and were much more sensitive to methemoglobinemia. Unfortunately these early observations were not followed up in more detail. During the past decade there has been a renewed interest on genetic polymorphism of drug oxidation, with the discovery of monogenic control in the oxidative metabolism of the antihypertensive drug debrisoquin (5) and sparteine, a drug with both oxytoxic and antiarrhythmic properties (6). Progress in this field has been enhanced by the discovery, identification, and characterization of different cytochrome P450 enzymes (CYP) (7). The main focus of this article is on the genetic polymorphism of CYP2D6. Another important enzyme example of genetic polymorphism of drug oxidation is that of the 4-hydroxylation of the *S*-enantiomer of the antiepileptic drug mephenytoin, which is the subject of a forthcoming review by the author.

Historical Perspective

An early study on the metabolic fate of debrisoquin at St. Mary's Hospital in London revealed the failure of a subject, unlike his colleagues, to oxidize the drug to its major 4-hydroxydebrisoquin metabolite (8). Years later, this same subject developed severe orthostatic hypotension within 2 hours of taking debrisoquin 32-mg base, a response that was not found in

---

This work was supported, in part, by the Pharmaceutical Research and Manufacturers of America Foundation and U.S. Public Health Service grant FD-T 000888.

From the Departments of Medicine and Pharmacology, Meharry Medical College, Nashville, Tennessee. Address reprint requests to Timi Edeki, M.D., Ph.D., Assistant Professor of Pharmacology and Medicine, Clinical Pharmacology Program, Meharry Medical College, 1005 Doctor D. B. Todd Jr. Boulevard, Nashville, TN 37208.

other subjects who ingested a similar dose (8). Subsequent analysis of the urine revealed that nonresponders to this dose excreted the drug as its major metabolite, 4-hydroxydebrisoquin, whereas the drug was excreted almost unchanged in the sensitive responder. These observations were followed up with a population study that led to the subsequent description of the genetic polymorphism of debrisoquin metabolism (5). At about the same time in Bonn, Germany, independent studies on sparteine showed unusual responses in some subjects. During kinetic studies with a slow-release preparation of sparteine (9), two subjects developed side effects such as diplopia, blurred vision, dizziness, and headache. These subjects had three to four times higher plasma levels of the sparteine, although all the subjects had the same dose. Further observations in these two subjects revealed that the two metabolites of sparteine, 2-dehydro- and 5-dehydrosparteine, usually present in plasma and urine, were not found in these subjects, and that nearly 100% of either an intravenous or oral dose was recovered in urine as sparteine. These initial observations and studies led to a population study with the subsequent description of another genetic drug polymorphism involving sparteine (6). The chemical structures of debrisoquin and sparteine and the main metabolites are shown in **Figure 1.**

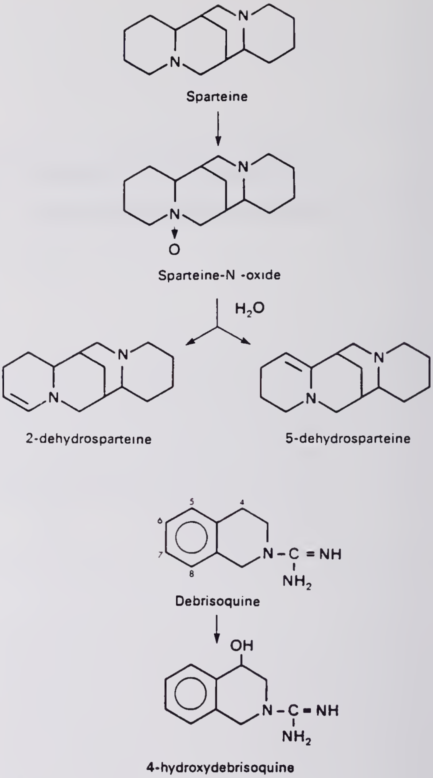


Fig. 1. Chemical structure of debrisoquin and sparteine and their main metabolites.

Methods Used in the Study of Drug Polymorphism

Several study methods are used in the identification of genetic polymorphism of drug metabolism to distinguish between hereditary and environmental factors contributing to variability in drug metabolism (10). These approaches are based on the use of carefully selected subjects under uniform and near basal environmental conditions.

Metabolic Phenotype. The first approach is the use of a specific probe drug to determine the metabolic phenotype of an individual. Pharmacokinetic parameters such as half-life, clearance, or metabolites of the drug might by measured in a large number of subjects. Based on the findings in these pharmacokinetic parameters, individuals are classified either as poor or extensive metabolizers (11, 12). An intermediate heterozygous group may also be evident.

Population Study. This allows an estimate of the frequency of a particular deficiency trait (e.g., metabolic capacity) and involves phenotyping a sufficient number of unrelated subjects from a homogenous population group (13). The frequency distribution of this deficiency may indicate whether the inheritance of this defect is either monogenic (Mendelian) or polygenic. The frequency distribution of a monogenic trait may be either bimodal or trimodal; but whereas if the trait polygenic, the frequency distribution is usually unimodal. The contribution of environmental factors in unimodal frequency distribution is often difficult to differentiate. Studies in different ethnic groups demonstrate the variation in the frequencies of the different phenotypes of drug oxidation in various populations **(Table 1).**

Twin Study. The relative contribution of genetic and environmental factors to drug metabolism is determined by investigations in groups of monozygotic (identical) and dizygotic (fraternal) twins under controlled clinical conditions. The contribution of heredity is calculated from a formula in which hereditability (H2) is the variance within pairs of dizygotic twins (VDz) minus variance within pairs of monozygotic twins (VMz), divided by the variance between pairs of dizygotic twins (14–15): $H2 = VDz - VMz/Vdz$, where variance (V) is calculated as the sum of (difference between twins)$^2/2N$ and N is the number of twin pairs of same zygosity. A value ranging from 0 (negligible genetic contribution) to 1 (virtually complete genetic influence) is possible. The twin study offers a convenient initial step; however, family studies are ultimately required.

Family/Pedigree Studies. The mode of inheritance is revealed by studying the transmission in families and generations of assigned phenotypes (16). If the mode of inheritance follows a classical mendelian pattern, the trait can be classified as autosomal dominant, autosomal recessive, or X-linked.

Panel Approach. Subjects who have been phenotyped with respect to a particular metabolic defect are given a single dose of the drug to be investigated and its pharmacokinetics and metabolism in these subjects are studied. These parameters are usually compared in two panels of subjects at both ends of the phenotype spectrum; for example, poor and extensive metabolizers. With this method, drugs whose metabolic patterns are affected by the genetic defect or polymorphism under investigation, are detected.

In Vitro and Enzyme Studies. In vitro techniques and enzyme studies are also used extensively as tools in the identification and study of genetic polymorphism of drug metabolism (17). Some of these techniques involve not just the use of liver microsomes, but also of isolated and purified enzymes and isoenzymes (18). *In vitro* techniques are useful in identifying the various forms of CYP involved in the oxidative metabolism of a particular drug and also to determine the potential for drug interactions (19, 20).

Genetic Studies. Application of the principles of molecular biology, such as polymerase chain reaction and restriction fragment length polymorphism, has been increasing in the study of genetic drug polymorphism (21). It has been demonstrated that poor metabolizers (PMs) of debrisoquin have negligible amounts of the CYP enzyme that metabolize this drug (22). PMs have mutant alleles that give rise to variant messenger RNA that produces a defective protein. DNA analysis of poor and extensive metabolizer subjects, have shown that different mutant alleles of varying combinations lead to different genotypes associated with both poor and extensive metabolizer

TABLE 1

Frequency of CYP2D6 Deficiency in Different Populations

Ethnicity (Nationality)	Probe Drug[a]	% (PMs/Total)	Reference
Africa			
Americans (USA, children)	DB	1.9 (2/106)	86
Ghanaians (Ghana)	SP	0 (0/154)	87
Ghanaians (Ghana)	DB	7 (10/14)	82
Ghanaians (Ghana)	DB	6.3 (5/80)	88
Nigerians (Nigeria)	SP	4 (7/165)	53
Nigerians (Nigeria)	SP	8 (10/123)	89
Nigerians (Nigeria)	DB	0 (0/138)	73
San Bushmen (South Africa)	DB	19 (18/96)	90
Zambians (Zambia)	DB	2 (2/102)	91
American Indians			
Cuna (Panama)	SP	0 (0/170)	92
Cuna (Panama)	DB	0 (0/89)	93
Ngawbe Guayme (Panama)	SP	5.2 (5/121)	92
Asia			
Chinese (China)	Metoprolol	0 (0/98)	94
Chinese (China)	DB	1 (7/695)	95
Filipino (Saudi Arabia)	DX	0 (0/55)	96
Japanese (Japan)	DB	0 (0/100)	97
Japanese (Japan)	SP	2 (2/84)	98
Japanese (Japan)	Metoprolol	0.5 (1/200)	94
Maori (New Zealand)	DB	5 (5/101)	99
Sinhalese (Sri Lanka)	DB	1.8 (2/111)	100
Thai (Thailand)	DB	1.2 (2/173)	101
Caucasians			
Americans (USA)	DX	6.7 (35/519)	102
Americans (USA)	DB	8.7 (16/183)	97
Americans (USA)	DB	7.7 (37/480)	86
Australians (Austria)	DB	6 (6/100)	103
Belgians (Belgium)	DB	7.2 (12/167)	104
Canadians (Canada)	SP	8.3 (4/48)	105
Danes (Denmark)	SP	7.3 (22/301)	33
Danes (Denmark)	SP	9.2 (33/358)	134
English (England)	DB	3 (3/94)	5
Estonians (Estonia)	DB	5.1 (4/78)	106
Estonians (Estonia)	DX	3.8 (3/78)	106
Finns (Finland)	DB	3.2 (5/155)	107
Lapps (Finland)	DB	8.6 (6/70)	107
French (France)	DX	7.4 (45/610)	108
French (France)	DX	3.9 (4/103)	109
French (France)	DX	3 (4/132)	110
French (France)	DX (ELISA)	5.1 (216)	111
Germans (Germany)	DX	10 (46/450)	112
Germans (Germany)	SP	7.7 (15/194)	113
Greenlanders (Greenland)	SP	3.2 (6/185)	114
West Greenlanders (Greenland)	SP	2.3 (4/171)	115
East Greenlanders (Greenland)	SP	3.3 (10/300)	115
Italians (Italy)	DX	4.5 (11/246)	116
Spanish (Spain)	DB	6.6 (25/377)	117
Spanish (Spain)	DB	11 (14/127)	118
Spanish (Spain)	DX	10 (15/146)	118
Swedes (Sweden)	DB	6.82 (69/1011)	95
Swedes (Sweden)	DB	5.4 (41/757)	63
Swedes (Sweden)	DB	8.8 (18/205)	119
Middle East			
Jordanians (Jordan)	DB	7.1 (14/195)	120
Jordanians (Jordan)	DB	7.7 (3/39)	120
Jordanians (Jordan)	DX	2.9 (7/241)	121
Saudis (Saudi Arabia)	DX	2 (2/102)	96
Saudis (Saudi Arabia)	DB	1 (1/102)	122
Turks (Turkey)	DB	3.4 (11/326)	123

[a] DB, debrisoquin; SP, sparteine; DX, dextromethorphan; ELISA, enzyme-linked immunosorbent-assay; PMs, poor metabolizers.

phenotypes (23). Thus, it is now possible by collection of a single blood sample to genotype individuals to determine whether they have metabolic deficiency traits (24).

Genetic Polymorphism of Debrisoquin/Sparteine/ Dextromethorphan Metabolism

Polymorphic hydroxylation of debrisoquin, an antihypertensive drug, was first described by Mahgoub et al. (5), and the polymorphic *N*-oxidation of sparteine was first described by Eichelbaum et al. (6). The metabolisms of these two drugs were later shown to be mediated by the same enzyme (25–27). Thus, if an individual is a PM of debrisoquin, he will also be a PM of sparteine. Another drug that has been used extensively in phenotyping individuals with respect to their hepatic drug-oxidizing capacity is dextromethorphan, an over-the-counter cough medication (28). It is metabolized to dextrorphan by cytochrome P4502D6 (CYP2D6). CYP2D6 is polymorphically distributed in different populations (Table 1) and is involved in the metabolism of debrisoquin and sparteine, which are also used as probe drugs, and more than 30 other clinically important drugs (29, 30). Based on their abilities to metabolize these drugs, individuals are classified as PMs or extensive metabolizers (EMs). The metabolic phenotype of a subject is assigned based on the metabolic ratio (MR) of the probe drug in that individual. The MR is determined as follows: MR = amount of debrisoquin (in moles)/ amount of 4-hydroxydebrisoquin (in moles) excreted in 8-hour urine; MR = amount of sparteine (in moles)/amount of 2,5-dehydrosparteine (in moles) excreted in 12-hour urine; MR = amount of dextromethorphan (in moles)/amount of dextrorphan (in moles) excreted in 8-hour urine. PMs are defined as having MR greater than 12.6, 20, or 0.3 for debrisoquin, sparteine, or dextromethorphan, respectively (27, 28, 31).

The genetic polymorphism of the metabolism of debrisoquin, sparteine, and dextromethorphan is not affected by age, sex, or smoking; it is transmitted in a Mendelian fashion and PM subjects are homozygous for an autosomal recessive gene (6, 16, 32–34). The frequency in the European population of the allele controlling poor metabolism is 0.2321 and that for extensive metabolism is 0.7679 (34). A study undertaken to find out if there is linkage of the polymorphic sparteine oxidation to various polymorphic marker systems showed a positive linkage to the P1 blood group (35). Because the P1 blood group has been mapped to the long arm of chromosome 22, it was concluded that the gene controlling debrisoquin/ sparteine metabolism polymorphism is situated on the longer arm of chromosome 22 in close vicinity to P1. No relationship has been found between debrisoquin/sparteine metabolism and drug acetylation or the excretion of bile acid and bile alcohol (36, 37), or 6-β-hydroxycortisol excretion (38). Furthermore, the phenotype of a subject is not affected by viral or pneumococcal vaccination (39) and it has been found to be stable after repeated phenotyping in patients with insulin-dependent diabetes mellitus (40).

Pharmacokinetics and Phenotype. The plasma concentrations of debrisoquin are significantly higher in PMs than in EMs (12). The 4-hydroxylation of debrisoquin by PM subjects after a 10-mg oral dose was capacity limited and dose dependent in the range of 1 to 20 mg, whereas this oxidation did not deviate from first-order kinetics over a dose of 10 to 40 mg in extensive metabolizers (41). PMs also tend to have increased pharmacodynamic response to drugs that are substrates of CYP2D6 (11). A study was carried out to examine the influence of various enzyme inducers, 1200 mg antipyrine, 100 mg phenobarbitone, 600 or 1200 mg rifampin daily for 7 days on sparteine metabolism and 6-β-hydrocortisol excretion in panels of EMs and PMs (42). Drug metabolism was induced in both EMs and PMs, as shown by increased 6-β-hydrocortisol excretion, although the increase caused by phenobarbitone was not significant, probably because of the low dose. Significant changes in elimination half-life and metabolic clearance were observed only in EMs after antipyrine and rifampin administration, with greater effect seen with 1200 mg/day rifampin. It was concluded that the regulation of cytochrome P-450 isozyme involved in debrisoquin/sparteine polymorphism is predominantly under genetic control, with only a marginal effect from enzyme induction.

Enantioselectivity and Debrisoquin 4-Hydroxylation. Two early reports have focused attention on the interesting phenomenon of enantioselectivity by debrisoquin-4-hydroxylation (43, 44). Debrisoquin has no chiral center; however, hydroxylation in position 4 leads to the formation of an asymmetric carbon center with two possible enantiomers with absolute configurations: *QR*(−)- and *S*(+)-4-hydroxydebrisoquin. EMs exhibited pronounced enantioselectivity in debrisoquin 4-hydroxylation with more than 99% of 4-hydroxydebrisoquin excreted as *S*(+)-enantiomer. The PMs, on the other hand, were characterized not only by reduced metabolite formation but also by loss of enantioselectivity, with 5 to 36% of the total 4-hydroxydebrisoquin excreted as the $R(-)$-enantiomer. The authors believed that 4-hydroxydebrisoquin enantiomer formation might provide an additional phenotyping criterion, especially in those subjects with metabolic ratios close to the antimode.

Substrates of CYP2D6. CYP2D6 is involved in the metabolism of more than 30 drugs that are very important in the clinical arena (29, 30) (see **Table 2**). Major classes include β-adrenoceptor blockers, antiarrhythmics, tricyclic antidepressants, neuroleptics, and serotonin reuptake inhibitors (20, 45).

Clinical and Therapeutic Implications

There are several important clinical, therapeutic, and toxicological implications of CYP2D6 deficiency.

Increased Side Effects. As a result of impaired metabolic capacity, PMs (even at normal doses) usually have higher plasma concentrations of drugs that are substrates of CYP2D6; for example, indoramin (46), amitriptyline, nortriptyline (47), timolol (12), and others (30). Thus, PM subjects may just be prone to the development of side effects during the course of a normal therapeutic use of a drug. Side effects in two PM subjects during the course of an initial investigation led to the discovery of the genetic polymorphism of sparteine metabolism (6). Therefore, to avoid increased side effects, such patients will require a lower-than-usual conventional dose (48), because their deficient metabolisms may lead to increased response (8, 11). The administration of prehexiline has been associated with peripheral neuropathy in debrisoquin-poor hydroxylation subjects (49). Phenformin 4-hydroxylation co-segregates with debrisoquin hydroxylation (50), and this might have been an important factor in the development of lactic acidosis when PM patients were given phenformin (51, 52). The metabolism of some β-adrenoceptor blockers such as bufuralol, timolol, propranolol, and metoprolol is controlled by the debrisoquin gene locus, although the metabolism of atenolol is not; thus PMs have increased plasma concentrations and may have increased β-adrenoceptor blockade and be prone to more side effects than EMs when given these drugs (53–61). Failure to consider the pharmacogenetic phenotype of a subject could therefore lead to over treatment of PMs and under treatment of EMs.

Drug Interactions. Certain drugs have very high affinity for CYP2D6, bind strongly to it, and prevent it from metabolizing other drugs. This includes CYP2D6 substrates such as propafenone, fluoxetine, and norfluoxetine, its main metabolite, and other drugs like quinidine, which, although it has high affinity for CYP2D6, is not metabolized by it (62). This inhibitory effect of quinidine has been demonstrated both *in vitro* and *in vivo*. Quinidine inhibits 2-hydroxylation of desipramine in both PMs and EMs, whereas its diastereoisomer quinine inhibits this metabolism

TABLE 2
Examples of CNS Drugs Metabolized by CYP2D6

Drug	Metabolic Route/Metabolite	Reference
Antidepressants		
Amitriptyline	Hydroxylation (Nortriptyline)	72
Clomipramine	2,8-Hydroxylation (hydroxyclomipramine)	124 125
Desmethylclomipramine	8-Hydroxylation (hydroxydesmethylclomipramine)	124
Desipramine	2-Hydroxylation (hydroxydesipramine)	126
Imipramine	2-Hydroxylation (hydroxylimipramine) (hydroxydesmethylimipramine)	127 128
Desmethylimipramine	2-Hydroxylation	70
Risperidone	9-Hydroxylation (hydroxyrisperidone)	129
Neuroleptics		
Reduced haloperidol	Oxidation (haloperidol)	130
Thioridazine	Side-chain oxidation (thioridazine 2-sulfoxide, thioradazine 2-sulfone)	72
Selective serotonin reuptake inhibitors		
Citalopram	Demethylation (desmethylcitalopram)	131
Desmethylcitalopram	Demethylation (didesmethylcitalopram)	131
Fluvoxamine	Hydroxylation	45
Paroxetine	Oxidation	78 45
Fluoxetine	Norfluoxetine	126
Sertraline	Demethylation (desmethylsertraline)	126
Nortriptyline	10-Hydroxylation (hydroxynortriptyline)	133
Chlorpropamide	2,3-Hydroxylation (hydroxychlorpropamide)	132

only in EMs, indicating a possible stereoselective mechanism (63). The high inhibitory potency of quinidine was recently demonstrated by the fact that even low doses of quinidine inhibited the metabolism of ophthalmic timolol, leading to significant increases in plasma timolol concentrations and pharmacodynamic response (12). *In vitro* studies have shown that the selective serotonin reuptake inhibitors paroxetine and fluvoxamine inhibit desipramine hydroxylation, albeit to different extents (45). Fluoxetine has been shown to reduce the oral clearance of tricyclic antidepressants (64, 65). Some clinical case reports have also provided evidence to suggest that fluoxetine interacts with tricyclic antidepressants and antipsychotics (66). The drug interactions between neuroleptics and tricyclic antidepressants (67) could also be explained on the basis of the involvement of CYP2D6. The potential for drug interaction is much greater in EMs than in PMs. It is also of interest that the metabolism of desipramine, a drug which co-segregates with debrisoquin metabolism (68–70), was inhibited by cimetidine in EMs but not in PMs (63). Because several drug types used in psychiatry are substrates of CYP2D6, pharmacogenetic evaluation is an important factor that must be considered in such clinical situations when psychotropic medications are administered. Studies have indicated that psychiatric drug treatment is capable of increasing the metabolic ratios and changing the pharmacogenetic status of patients (71–73).

Shift to a More Toxic Metabolite. Another important consequence of impaired drug metabolism is a shift in the metabolic pathway leading to the formation of a toxic metabolite, when the main pathway of detoxication is impaired, while that of the toxic metabolite is not. This is exemplified by phenacetin, in which its impaired *O*-deethylation leads to the formation of increased 2-hydroxyphenetidin, a metabolite implicated in the development of methemoglobinemia. This was confirmed by more cases of methemoglobinemia in PMs than in EMs after administration of the same dose of phenacetin (74).

Therapeutic Failure. Defective drug metabolism could also lead to therapeutic failure, if impaired formation of an active metabolite is responsible for therapeutic activity. An example of this is the antiarrhythmic drug encainide, in which the *O*-demethylated metabolites have antiarrhythmic activity and their formation shares genetic control similar to that of debrisoquin (75, 76). Another example of therapeutic failure occurring in poor metabolizers is codeine, which is *O*-demethylated to morphine by CYP2D6 (77). The role of the genetic phenotype of a subject in the analgesic efficacy of codeine is supported by a study in which codeine increased the pain threshold in EMs but not in PMs (78), thus providing evidence that morphine formation is essential for analgesia during codeine treatment. In comparison with EMs, PMs may therefore not benefit as much from the analgesic effect of codeine.

Rapid Metabolism Leading to Therapeutic Failure. The phenomenon of "ultra rapid metabolism" of debrisoquin was first revealed by the case of a depressed patient who was taking nortriptyline and who failed to attain adequate therapeutic plasma concentrations because of very rapid metabolism of nortriptyline (79). As a result of restriction fragment length polymorphism analysis of the CYP2D locus in two families, the cause of the very rapid metabolism of debrisoquin was identified as being a variant CYP2D6 gene, CYP2D6L (80); the members of these families had inherited amplification of the CYP2D6. The prevalence of subjects with duplication of CYP2D6 in a Spanish population was about 7% (81), similar to that of poor metabolizers. Ultra rapid metabolism, therefore, is another important factor that must be considered to avoid therapeutic failure of CYP2D6 substrates.

Interethnic Differences. The frequency of the PM phenotypes varies in populations with different ethnic origins (Table 1). Most of the studies on deficient metabolism of drugs by poor metabolizers have been carried out in Caucasian populations. However, these findings may not always apply to other populations; for instance, one study failed to show that poor metabolizers of sparteine are also poor metabolizers of debrisoquin, quite unlike the findings in Caucasian populations (82). A study in a Nigerian population did not identify poor metabolizers of metoprolol or debrisoquin. Differences in gene frequencies have also been demonstrated in various populations. Spaniards had a lower frequency of the CYP2D6(B) allele and a higher frequency of the wild-type allele than other Caucasian populations (83). The CYP2D6*Ch* alleles were very frequent in Orientals populations, whereas, CYP2D6A or CYP2D6B were not identified (84). Further, the lower prevalence of debrisoquin poor metabolizers was said to be caused by reduced frequency of the CYP2D6(B) mutation (85).

Implications for Drug Development. Genetic polymorphism of drug oxidation has implications in the development of new drugs (10, 27). First, subjects should be enlisted for the initial phase I and II studies of new drugs based on their phenotypes with probe drugs that have already been established. This will avoid the use of subjects from one extreme of the phenotype spectrum. Second, the varying frequency of polymorphism phenotypes in different ethnic groups raises doubts about the reliability of extrapolating findings of drug studies in one population group to others. To give reliable dosage recommendations in particular populations, studies will need to conducted using subjects representative of that particular population. Extensive application of *in vitro* drug metabolism studies has provided early information on the various types of CYP involved in the metabolism of a particular drug and on potential drug interactions. This has enabled more meaningful clinical studies to be conducted in humans.

Conclusion

The recent advances in the field of genetic polymorphism of drug oxidation has revealed the important concept that pharmacogenetic evaluation should be considered in drug therapy of patients. This will not only allow optimization of drug therapy, but also avoid unnecessary side effects of drugs. These considerations are especially relevant in the use of cardiovascular drugs and in the treatment of psychiatric disorders.

References

1. Chapman CJ. Drugs and genes. Therapeutic implications. Drugs 1977; 14:120–127.
2. Johansson I, Ludqvist E, Bertilsson L, et al. Inherited amplification of an active gene in the cytochrome P450 CYP2D locus as a cause of ultrarapid metabolism of debrisoquine. Proc Natl Acad Sci USA 1993; 90: 11825–11829.
3. Edeki T, Brase D. Phenytoin disposition and toxicity. Role of pharmacogenetic and interethnic factors. Drug Metab Rev 1995; 27:449–469.
4. Shahidi NT. Acetophenetidin-induced methemoglobinemia. Ann NY Acad Sci 1968; 151:822–832.
5. Mahgoub A, Idle JR, Dring LG, Lancaster R. Polymorphic hydroxylation of debrisoquine in man. Lancet 1977; 2:584–586.
6. Eichelbaum M, Spannbrucker N, Steinecker B, Dengler HJ. Defective N-oxidation of sparteine in man. A new pharmacogenetic defect. Eur J Clin Pharmacol 1979; 16:183–187.
7. Guengerich FP. Human cytochrome P-450 enzymes. Life Sci 1992; 0:1471–1478.
8. Smith RL. Discovery of the genetic polymorphism of drug oxidation. Xenobiotica 1986; 16:361–365.
9. Eichelbaum M, Spannbrucker N, Dengler HJ. Lack of N-oxidation of sparteine in certain healthy subjects. Sixth International Congress of Pharmacology; 1975 July 20–25; Helsinki. 1975: 1071.
10. Eichelbaum M. Polymorphism of drug oxidation in man: novel findings. Trends Pharmacol Sci 1981.
11. Al-Sereiti MD, Edeki T, Lledo P, Turner P. The effects of timolol on intraocular pressure and exercise heart rate in poor and extensive debrisoquine metabolizers. Int J Clin Pharmacol Res 1990; 6:339–345.
12. Edeki TI, He H, Wood AJJ. Pharmacogenetic explanation for excessive B-blockade following timolol eye drops. Potential for oral-ophthalmic drug interaction. J Am Med Assoc 1995; 274:1611–1613.
13. He N, Edeki T. Polymorphism of dextromethorphan oxidation in African Americans [Abstract]. J Invest Med 1995; 43:256A.
14. Woosely RL, Roden DM, Duff HJ, et al. Coinheritance of deficient oxidative metabolism of encainide and debrisoquine. Clin Res 1981; 29:501A.
15. Penno MB, Dvochik BH, Vesell ES. Genetic variation in rates of antipyrine metabolism formation. A study in uninduced twins. Proc Natl Acad Sci USA 1981; 78: 5193–5196.
16. Evans DAP, Harmer D, Downham DY, et al. The genetic control of sparteine and debrisoquine metabolism in man with new methods of analyzing bimodal distributions. J Med Genet 1983; 20:321–329.
17. Oates NS, Shah RR, Idle JR, Smith RI. Genetic polymorphism of phenformin 4-hydroxylation. Clin Pharmacol 1982; 32:81–89.
18. Distlerath LM, Reilly PEB, Martin AV, et al. Purification and characterization of the human liver cytochromes P-450 involved in debrisoquine 4-hydroxylation and phenacetin *o*-deethylation, two prototypes for genetic polymorphism in oxidative drug metabolism. J Biol Chem 1985; 260:9057–9067.
19. Otton SV, Wu D, Joffe RT, et al. Inhibition by fluoxetine of cytochrome P450 2D6 activity. Clin Pharmacol Ther 1992; 53:401–409.
20. Otton V, Schadel M, Cheung S, et al. CYP2D6 phenotype determines the metabolic conversion of hydrocodone to hydromorphone. Clin Pharmacol Ther 1993; 54:463–472.
21. Steen VM, Andreassen OA, Daly AK, et al. Detection of the metabolizer-associated CYP2D6(D) gene deletion allele by long-PCR technology. Pharmacogenetics 1995; 5:215–223.
22. Gonzalez FJ, Skoda CR, Kimura SM, et al. Characterization of the common genetic defect in humans deficient in debrisoquine metabolism. Nature 1988; 331: 442–446.
23. Skoda CK, Gonzalez FJ, Demierre A, Meyer UA. Two mutant alleles of the human cytochrome P-450 dbl gene (P450C2D1) associated with genetically deficient metabolism of debrisoquine and other drug. Proc Natl Acad Sci USA 1988; 85:5240–5243.
24. Broly F, Gaedigk A, Heim M, et al. Debrisoquine sparteine hydroxylation genotype and phenotype. Analysis of common mutations and alleles of CYP2D6 in a European population. DNA Cell Biol 1991; 10: 545–558.
25. Inaba T, Otton SV, Kalow W. Deficient metabolism of debrisoquine and sparteine. Clin Pharmacol Ther 1980; 27:547–549.
26. Inaba T, Vinks A, Otton SV, Kalow W. Comparative pharmacogenetics of sparteine and debrisoquine. Clin Pharmacol Ther 1983; 33:394–399.
27. Eichelbaum M, Bertilsson L, Sawe J, Zekorn C. Polymorphic oxidation of sparteine and debrisoquine related pharmacogenetic entities. Clin Pharmacol Ther 1982; 31:184–186.
28. Schmid B, Bircher J, Preisig R, Kupfer A. Polymorphic dextromethorphan metabolism. Cosegregation of oxidative *O*-demethylation with debrisoquin hydroxylation. Clin Pharmacol Ther 1985; 38:618–624.

29. Jacqz E, Hall SD, Branch RA. Genetically determined polymorphisms in drug oxidation. Hepatology 1986; 6:1020–1032.
30. Eichelbaum M, Gross AS. The genetic polymorphism of debrisoquine/sparteine metabolism. Clin Asp Pharmacol Ther 1990; 46:377–394.
31. Evans DAP, Maghoub A, Sloan TP, et al. A family and population study of genetic polymorphism of debrisoquine oxidation in a white British population. J Med Genet 1980; 17:102–105.
32. Idle J, Smith RL. The debrisoquine hydroxylation gene. A gene of multiple consequence. Am Soc Pharmacol Exp Ther 1984; 80:48–164.
33. Brosen K, Otton SV, Gram LF. Sparteine oxidation polymorphism in Denmark. Acta Pharmacol Toxicol 1985; 57:357–360.
34. Eichelbaum M, Reetz K-P, Schmidt EK, Zekorn C. The genetic polymorphism of sparteine metabolism. Xenobiotica 1986; 16:405–481.
35. Eichelbaum M, Baur MP, Dengler HJ, et al. Chromosome assignment of human cytochrome P-450 (debrisoquine/sparteine type) to chromosome 22. Br J Clin Pharmacol 1987; 23:455–458.
36. Karlaganis G, Kupfer A, Preisig R. Urinary bile alcohol excretion does not reflect the genetic polymorphism of debrisoquine hydroxylation. Br J Clin Pharmacol 1984; 17:470–473.
37. Harmer D, Pevans DA, Ezc LC, et al. The relationship between the acetylator and the sparteine hydroxylation polymorphisms. J Med Genet 1986; 23:155–156.
38. Park BK, Eichelbaum M, Ohnhaus EE. 6-beta hydroxycortisol excretion in relation to polymorphic N-oxidation of sparteine [Letter]. Br J Clin Pharmacol 1982; 13:737–740.
39. Ayesh R, Scandding G, Brostoff J, et al. Stability of the debrisoquine metabolic ratio to immunoperturbation by influenza and pneumococcus vaccines [Abstract]. Br J Clin Pharmacol 1988; 25:141P.
40. Bechtel YC, Joanne C, Grandmottet M, Bechtel PR. The influence of insulin-dependent diabetes on the metabolism of caffeine and the expression of debrisoquine. Clin Pharmacol Ther 1988; 44:408–417.
41. Sloan TP, Lancaster R, Shann RR, et al. Genetically determined oxidation capacity and the disposition of debrisoquine. Br J Clin Pharmacol 1983; 5:443–450.
42. Eichelbaum M, Mineshita S, Ohnshaus EE, Zekorn C. The influence of enzyme induction Hon polymorphism sparteine oxidation. Br J Clin Pharmacol 1986; 22:49–53.
43. Meese CO, Bertilsson L, Fisher C, et al. Stereoselectivity of 4-hydroxydebrisoquine enantiomer formation in relation to debrisoquine phenotype. Presented at Third World Conference on Clinical Pharmacology and Therapeutics; July to August 1986.
44. Eichelbaum M. Defective oxidation of drugs. Pharmacokinetic and therapeutic implications. Clin Pharmacokinet 1982; 7:1–22.
45. von Moltke L, Greenblatt D, Court M, et al. Inhibition of alprazolam and desipramine hydroxylation in vitro by paroxetine and fluvoxamine. Comparison with other selective serotonin reuptake inhibitor antidepressants. J Clin Psychopharmacol 1995; 15:125–131.
46. Pierce DM, Smith SE, Franklin RA. The pharmacokinetics of Indoramin and 6-hydroxyindoramin on poor and extensive hydroxylators of debrisoquine. Eur J Clin Pharmacol 1987; 33:59–65.
47. Baumann P, Jonzier-Percy M, Koeb L, et al. Amitriptyline pharmacokinetics and clinical response: 11. Metabolic polymorphism assessed by hydroxylation of debrisoquine and mephenytoin. Int Clin Psychopharmacol 1986; 1:102–112.
48. Silas JH, Lennard MS, Tucker GT, et al. Why hypertensive patients vary their response to oral debrisoquine. Br Med J 1977; 1:422–425.
49. Shah RR, Oates NS, Idle JR, et al. Prediction of subclinical prehexiline neuropathy in a patient with inborn error of debrisoquine hydroxylation. Am Heart J 1988; 105:159–161.
50. Oates NS, Shah RR, Idle JR, Smith RI. Genetic polymorphism of phenformin 4-hydroxylation. Clin Pharmacol 1982; 32:81–89.
51. Shah RR, Oates NS, Idle JR, Smith RL. Genetic impairment of phenformin metabolism. Lancet 1980; 1:1147.
52. Wiholm B, Alvan G, Bertilsson L, et al. Hydroxylation of debrisoquine in patients with lactic acidosis after phenformin. Lancet 1981; 1:1098–1099.
53. Lennard MS, Silas JH, Freestone S, et al. Oxidation phenotype: A major determinant of metoprolol metabolism and response. N Engl J Med 1982; 307:1558–1560.
54. Lennard MS, Jackson PR, Freestone S, et al. The relationship between debrisoquine oxidation phenotype and the pharmacokinetics and pharmacodynamics of propranolol. Br J Clin Pharmacol 1982; 17:679–685.
55. Lennard MS, Tucker GT, Silas JH, Woods HF. Debrisoquine polymorphism and the metabolism and action of metoprolol, timolol, propranolol, and atenolol. Xenobiotica 1986; 16:435–447.
56. Lennard MS, Iyun AO, Jackson PR, et al. Evidence for a dissociation in the control of sparteine, debrisoquine and metoprolol metabolism in Nigerians. Pharmacogenetics 1992; 2:89–92.
57. Dayer P, Kubli A, Kupfer A, et al. Defective hydroxylation of bufuralol associated with side-effects of the drug in poor metabolisers. Br J Clin Pharmacol 1982; 13:750–751.
58. Dayer P, Balant L, Kupfer A, et al. Contribution of the genetic status of oxidation metabolism to variability in the plasma concentrations of β-adrenoceptor blocking agents. Eur J Clin Pharmacol 1983; 24:797–799.
59. Raghuram TC, Koshakji RP Wilkinson GR, Wood AJJ. Polymorphic ability to metabolize propranolol alters 4-hydroxypropranolol levels but not beta blockade. Clin Pharmacol Ther 1984; 36:51–56.
60. McGourty JC, Silas JH. Metoprolol metabolism and debrisoquine oxidation polymorphism: Population and family studies. Br J Clin Pharmacol 1985; 20:555–566.
61. McGourty JC, Silas JH, Fleming JJ, et al. Pharmacokinetics and beta-blocking effects of timolol in poor and extensive metabolisers of debrisoquin. Clin Pharmacol Ther 1985; 38:409–413.
62. Mikus G, Ha HR, Vozen S, et al. Pharmacokinetics and metabolism of quinidine in extensive and poor metabolisers of sparteine. Eur J Clin Pharmacol 1986; 31: 69–72.
63. Steiner E, Bertilsson L, Sawe J, et al. Polymorphic debrisoquin hydroxylation in 757 Swedish subjects. Clin Pharmacol Ther 1988; 431–435.
64. Bergstrom RF, Peyton AL, Lemberger L. Quantification and mechanism of the fluoxetine and tricyclic antidepressant interaction. Clin Pharmacol Ther 1992; 51: 239–248.

65. Preskorn SH, Alderman J, Chung M, et al. Pharmacokinetics of desipramine coadministered with sertraline or fluoxetine. J Clin Psychopharmacol 1994; 14: 90–98.
66. Ciraulo DA, Shader RI. Fluoxetine drug-drug interactions: 1. Antidepressants and antipsychotics. J Clin Psychopharmacol 1990; 10:48–50.
67. Gram LF, Overo KE. Drug interaction. Inhibitory effect of neuroleptics on metabolism of tricyclic antidepressants in man. Br Med J 1972; 1:463–465.
68. Bertilsson L, Aberg-Wistedt A. The debrisoquine hydroxylation text predicts steady-state plasma levels of desipramine. Br J Clin Pharmacol 1983; 15:388–390.
69. Spina E, Birgersson C, von Bahr C, et al. Phenotypic consistency in hydroxylation of desmethylimipramine and debrisoquine in healthy subjects and in human liver microsomes. Clin Pharmacol Ther 1984; 36:677–682.
70. Spina E, Steiner E, Ericsson O, Sjoqvist F, Hydroxylation of desmethylimipramine. Dependence on debrisoquine hydroxylation phenotype. Clin Pharmacol Ther 1987; 41:314–319.
71. Derenne F, Joanne C, Vandel S, et al. Debrisoquine oxidative phenotyping and psychiatric drug treatment. Eur J Clin Pharmacol 1989; 36:53–58.
72. Baumann P, Meyer JW, Amey M, et al. Dextromethorphan and mephenytoin phenotyping of patients treated with thioridazine or amitriptyline. Ther Drug Monit 1992; 14:1–8.
73. Llerena A, Herraiz AG, Cobaleda J, et al. Debrisoquin and mephenytoin hydroxylation phenotypes and CYP2D6 genotype in patients treated with neuroleptic and antidepressant agents. Clin Pharmacol 1993; 54: 606–611.
74. Ritchie JC, Sloan TP, Idle JP, Smith RL. Toxicological implications of polymorphic drug metabolism. Ciba Found Symp 1980; 76:219–244.
75. Woosely RL, Roden DM, Duff HJ, et al. Coinheritance of deficient oxidative metabolism of encainide and debrisoquine. Clin Res 1981; 29:501A.
76. McAllister CB, Wolfenden HT, Aslanian WS, et al. Oxidative metabolism of encainide: polymorphism, pharmacokinetics and clinical considerations. Xenobiotica 1986; 16:483–490.
77. Yue QY, Svensson JO, Alm C, et al. Codeine O-demethylation co-segregates with polymorphic debrisoquine hydroxylation. Br Clin Pharmacol 1989; 28:639–645.
78. Sindrup SH, Brosen K, Bjerring P, et al. Codeine increases pain thresholds to copper vapor laser stimuli in extensive but not poor metabolizers of sparteine. Clin Pharmacol Ther 1990; 48:686–693.
79. Bertilsson L, Aberg-Wistedt A, Gustafsson LL, Nordin C. Extremely rapid hydroxylation of debrisoquine. A case report with implication for treatment with nortriptyline and other tricyclic antidepressants. Ther Drug Monit 1985; 7:478–480.
80. Johansson I, Ludqvist E, Bertilsson L, et al. Inherited application of an active gene in the cytochrome P450 CYP2D locus as a cause of ultrarapid metabolism of debrisoquine. Proc Natl Acad Sci USA 1993; 90: 11825–11829.
81. Agundez J, Ledesma MC, Ladero JM Benitez J. Prevalence of CYP2D6 gene duplication and its repercussion on the oxidative phenotype in a white population. Clin Pharmacol Ther 1995; 57:265–269.
82. Woolhouse NM, Eichelbaum M, Oates NS, et al. Dissociation of co-regulatory control of debrisoquin/phenformin and sparteine oxidation in Ghanaians. Clin Pharmacol Ther 1985; 37:512–521.
83. Agundez J, Martinez C, Ledesma MC, et al. Genetic basis for differences in debrisoquin polymorphism between a Spanish and other white populations. Clin Pharmacol Ther 1994; 55:412–417.
84. Dayer P, Kubli A, Kupfer A, et al. Defective hydroxylation of bufuralol associated with side-effects of the drug in poor metabolisers. Br J Clin Pharmacol 1982; 13:750–751.
85. Evans WE, Relling MV, Rahman A, et al. Genetic basis for a lower prevalence of deficient CYP2D6 oxidative metabolism phenotypes in Black Americans. J Clin Invest 1993; 91:2150–2154.
86. Relling MV, Cherrie J, Schell MJ, et al. Lower prevalence of the debrisoquin oxidative poor metabolizer phenotype in American black versus white subjects. Clin Pharmacol Ther 1991; 50:308–313.
87. Eichelbaum M, Baur MP, Dengler HJ, et al. Chromosome assignment of human cytochrome P-450 (debrisoquine/sparteine type) to chromosome 22. Br J Clin Pharmacol 1987; 23:455–458.
88. Woolhouse NM, Andoh B, Mahgoub A, et al. Debrisoquine hydroxylation polymorphism among Ghanaians and Caucasians. Clin Pharmacol Ther 1979; 26:584–591.
89. Mbanefo C, Bababunmi EA, Mahgoub A, et al. A study of the debrisoquine hydroxylation polymorphism in a Nigerian population. Xenobiotica 1980; 10:811–818.
90. Sommers DK, Monerieff J, Avenant J. Polymorphism of the 4-hydroxylation of debrisoquine in the San Bushmen of Southern Africa. Hum Toxicol 1988; 7:273–276.
91. Simooya O, Njunju E, Hodjegan AR, et al. Debrisoquine and metoprolol oxidation in Zambians: a population study. Pharmacogenetics 1993; 3:205–208.
92. Arias TD, Inaba T, Cooke RG, Jorge LF. A preliminary note on the transient polymorphic oxidation of sparteine in the Ngawbe Guaymi Amerindians. A case of genetic divergence with tentative phylogenetic time from for the pathway. Clin Pharmacol Ther 1988; 44: 343–352.
93. Jorge LF, Arias TD, Inaba T, Jackson PR. Unimodal distribution of the metabolic ratio for debrisoquine in Cuna Amerindians of Panama. Br J Clin Pharmacol 1990; 30:281–285.
94. Horai Y, Nakano M, Ishizaki T, et al. Metoprolol and mephenytoin oxidation polymorphisms in Far Eastern Oriental subjects. Japanese versus mainland Chinese. Clin Pharmacol Ther 1989; 46:198–207.
95. Bertilsson L, Lou YQ, Yun-Long D, et al. Pronounced differences between native Chinese and Swedish populations in the polymorphic hydroxylations of debrisoquin and S-mephenytoin. Clin Pharmacol Ther 1992; 51:388–397.
96. Evans DA, Krahn P, Narayanan N. The mephenytion (cytochrome P-450 2C 19) and dextromethorphan (cytochrome P450 2D6) polymorphisms in Saudi Arabians and Filipinos. Pharmacogenetics 1995; 5:64–71.
97. Nakamura K, Goto F, Ray WA, et al. Interethnic differences in genetic polymorphism of debrisoquin and mephenytoin hydroxylation between Japanese and Caucasian populations. Clin Pharmacol Ther 1985; 10: 402–408.
98. Ishizaki T, Eichelbaum M, Horai Y, et al. Evidence for polymorphic oxidation of sparteine in Japanese subjects. Br J Clin Pharmacol 1987; 23:482–485.

99. Wanwimolruk S, Pratt EL, Denton JR, et al. Evidence for the polymorphic oxidation of debrisoquine and proguanil in a New Zealand Maori population. Pharmacokinetics 1995; 5:193–198.
100. Weerasuriya K, Jayakody RL, Smith AD, et al. Debrisoquine and mephenytoin oxidation in Sinhalese: a population study. Br J Clin Pharmacol 1994; 38:466–470.
101. Wanwimolruk S, Patamasucon P, Lee EFD. Evidence for the polymorphic oxidation of debrisoquine in the Thai population. Br J Clin Pharmacol 1990; 29:244–247.
102. Guttendorf RJ, Britto M, Blouin RA, et al. Rapid screening for polymorphisms in dextromethorphan and mephenytoin metabolism. Br J Clin Pharmacol 1990; 29: 373–380.
103. Peart GF, Boutagy J, Shenfield GM. Debrisoquine oxidation in an Australian population. Br J Clin Pharmacol 1986; 21:465–471.
104. Leclercq V, Desager JP, Nieuwenhuyze YV, Harvengt C. Prevalence of drug hydroxylator phenotypes in Belgium. Eur J Clin Pharmacol 1987; 33:439–440.
105. Vinks A, Inaba T, Otton SV, Kalow W. Sparteine metabolism in Canadian Caucasians. Clin Pharmacol Ther 1982; 31:23–29.
106. Kuvet RA, Svenson JO, Bertilsson L, Sjoqvist. Polymorphism of debrisoquine and mephenytion hydroxylation among Estonians. Pharmacol Toxicol 1993; 72:113–115.
107. Arvela P, Kirjarinta M, Karki N, Pelkonen O. Polymorphism of debrisoquine hydroxylation among Finns and Lapps. Br J Clin Pharmacol 1988; 26:601–603.
108. Viry MV, Fournier B, Siest G, Galteau MM. Dextromethorphan O-demethylation in a large number of French Caucasian families. Pharmacogenetics 1992; 2:135–138.
109. Larrey D, Amouyal G, Tinel M, et al. Polymorphism of dextromethorphan oxidation in a French population. Br J Clin Pharmacol 1987; 24:676–679.
110. Jacqz E, Dulae H, Mathieu H. Phenotyping polymorphic drug metabolism in the French Caucasian population. Eur J Clin Pharmacol 1988; 35:167–171.
111. Freche JP, Dragacci SM, Petit AM, et al. Development of an ELISA to study the polymorphism of dextromethorphan oxidation in a french population. Eur J Clin Pharmacol 1990; 39:481–485.
112. Hildebrand M, Seifert W, Reichenberger. Determination of dextromethorphan metabolizer phenotype in healthy volunteers. Eur J Clin Pharmacol 1989; 36: 315–318.
113. Bock KW, Schrenk D, Forster A, et al. The influence of environmental and genetic factors on CYP2D6, CYP1A2 and UDP-glucuronosyltransferases in man using sparteine, caffeine, and paracetamol as probes. Pharmacogenetics 1994; 4:209–218.
114. Brosen K. Sparteine oxidation polymorphism in Greenlanders living in Denmark. Br J Clin Pharmacol 1986; 22:415–419.
115. Clasen K, Madsen L, Brosen K, et al. Sparteine and mephenytoin oxidation. Genetic polymorphisms in East and West Greenland. Clin Pharmacol Ther 1991; 49: 624–631.
116. Spina E, Campo GM, Avenoso A, et al. CYP2D6 related oxidation polymorphism in Italy. Pharmacol Res 1994; 29:281–289.
117. Benitez J, Llerena A, Cobaleda J. Debrisoquin oxidation polymorphism is a Spanish population. Clin Pharmacol Ther 1988; 44:74–77.
118. Henthorn TK, Benitez J, Avram MJ, et al. Assessment of the debrisoquin and dextromethorphan phenotyping tests by gaussian mixture distributions analysis. Clin Pharmacol Ther 1989; 45:328–333.
119. Sanz EJ, Villen T, Alm C, Bertilsson L. S-mephenytoin hydroxylation phenotypes in a Swedish population determined after coadministration with debrisoquin. Clin Pharmacol Ther 1989; 45:495–498.
120. Al-Hadidi HF, Irshaid YM, Rawashdeh NM. Metoprolol α-hydroxylation is a poor probe for debrisoquin oxidation (CYP2D6) polymorphism in Jordanians. Eur J Clin Pharmacol 1994; 47:311–314.
121. Irshaid YM, Al-Hadidi HF, Rawashdeh NM. Dextromethorphan O-demethylation polymorphism in Jordanians. Eur J Clin Pharmacol 1993; 45:271–273.
122. Islam SI, Idle JR, Smith RL. The polymorphic 4-hydroxylation of debrisoquine in a Saudi-Arab population. Xenobiotica 1980; 10:819–825.
123. Bozkurt A, Basci N, Isimer A, et al. Polymorphic debrisoquin metabolism in a Turkish population. Clin Pharmacol Ther 1994; 55:399–401.
124. Gorgia AE, Baint LP, Genet C, et al. Importance of oxidative polymorphism and levomepromazine treatment on the steady-state blood concentrations of clomipramine and its major metabolites. Eur J Clin Pharmacol 1986; 31:449–455.
125. Nielson K, Brosen K, Hansen M, Gram L. Single-dose kinetics of clomipramine. Relationship to the sparteine and S-mephenytoin oxidation polymorphisms. Clin Pharmacol Ther 1994; 55:518–527.
126. von Molte L, Greenblatt D, Coteau-Bibbo M, et al. Inhibition of desipramine hydroxylation in vitro by serotonin reuptake inhibitor antidepressants, and quinidine and ketoconazole. A model system to predict drug interacts in vivo. J Pharmacol Exp Ther 1994; 268: 1278–1283.
127. Brosen K, Gram L. Quinidine inhibits the 2-hydroxylation of imipramine and desipramine but not the demethylation of imipramine. Eur J Clin Pharmacol 1987; 37:155–160.
128. Brosen K, Zeugin T, Meyer U. Role of P450IID6, the target of the sparteine-debrisoquin oxidation polymorphism, in the metabolism of imipramine. Clin Pharmacol Ther 1991; 49:609–617.
129. Huang ML, Peer AV, Woestenborghs R, et al. Pharmacokinetics of the novel antipsychotic agent risperidone and the prolactin response in healthy subjects. Clin Pharmacol Ther 1993; 54:257–268.
130. Llerena A, Dahl M, Ekqvist B, Bertilsson L. Haloperidol disposition is dependent on the debrisoquine hydroxylation phenotype. Increased plasma levels of the reduced metabolite in poor metabolizers. Ther Drug Monit 1992; 14:261–264.
131. Sindrup SH, Brosen K, Hansen MG, et al. Pharmacokinetics of citalopram in relation to the sparteine and the mephenytoin oxidation polymorphisms. Ther Drug Monit 1993; 15:11–17.
132. Kallio J, Huupponeen R, Pyykko K. The relationship between debrisoquine oxidation phenotype and the pharmacokinetics of chlorpropamide. Eur J Clin Pharmacol 1990; 39:93–95.
133. Woodhouse N, Yamoah K, Mellstrom B, et al. Nortriptyline and debrisoquine hydroxylations in Ghanian and Swedish subjects. Clin Pharmacol Ther 1984; 36: 374–378.
134. Drohse A, Bathum L, Brogen K, Gram LF. Mephenytoin and sparteine oxidation: genetic polymorphisms in Denmark. Br J Clin Pharmacol 1989; 27:620–625.

The Art and Science of the Psychopharmacotherapy of African Americans

William B. Lawson, M.D., Ph.D.

Abstract

Recent research and clinical experience has shown that African Americans may be at greater risk for inappropriate treatment. Such experiences can interact negatively with an existing distrust of the mental health system. Providers may show different prescribing patterns with racial and ethnic minorities: they may overuse antipsychotics, dispense higher dosages, and more commonly give involuntary treatment, which results in more side effects and a poorer outcome. Conversely, they may underuse other psychotropic medications, especially for anxiety and affective disorders, which are underdiagnosed in minorities. Recent research suggests that ethnic differences may exist in pharmacokinetics, and so different dosing strategies may be necessary. Not surprisingly African Americans in distress are more likely to seek initial treatment outside of the mental health system, seek treatment later in the course of the illness, complain more about side effects, and terminate treatment earlier. Cultural as well as socioeconomic factors must be considered. Newer pharmacological agents may be potentially more helpful for minorities because they are better tolerated, have better side effect profiles, and demonstrate better efficacy. However, African Americans have limited access to these agents. Education of providers and patients, policy changes in the public sector, wider implementation of research policies concerning inclusion of minorities, and different marketing strategies by pharmaceutical concerns are probably necessary to maximize pharmacotherapy of minorities.

HISTORICALLY, RACIAL DIFFERENCES have been reported in psychiatric diagnostic categories. African Americans were more likely to receive the clinical diagnosis of schizophrenia than the diagnosis of an affective disorder (1, 2). These differences may be either a consequence of "true" differences in prevalence, differences in phenomenology rather than prevalence (i.e., African Americans with the same disorder as Caucasians may have different presenting symptoms), or these differences may be a consequence of misdiagnosis (3). When structured interviews are used, racial differences tend to disappear (2). Moreover, large-scale studies such as the Epidemiological Catchment Area study, a five-city door-to-door survey using a structured interview to determine diagnoses, and the National Comorbidity Study, a recent national randomized survey using a structured interview that could generate DSM IIIR (Diagnostic and Statistical Manual of Mental Disorders, 3rd edition, revised) diagnoses, found few consistent racial or ethnic differences (2). The National Comorbidity Study tended to show that minorities had a lower prevalence of most mental disorders, including schizophrenia, than nonhispanic whites (4). However that study did not include institutional settings in which minorities are overrepresented. Such findings suggest that previously reported racial differences

---

From the North Little Rock Veterans Administration Medical Center, North Little Rock, Arkansas. Address reprint requests to William B. Lawson, M.D., Ph.D., Chief of the Chronically Mentally Ill, 116A NLR VAMC, 2200 Fort Roots Drive, North Little Rock, AR 72114.

in diagnoses were largely caused by diagnostic error.

Racial misdiagnosis could lead to inappropriate treatment in African Americans, including the use of the wrong medication, which could lead to a poorer outcome or exclusion of needed medication. Often African Americans are overdiagnosed as having schizophrenia and may be inappropriately receiving antipsychotic medication (5). Bell and Mehta (6, 7) reported that African American patients with lithium-responsive bipolar affective disorder and excellent lithium response often received a diagnosis of schizophrenia and were treated with antipsychotic drugs rather than lithium. We reported that schizophrenia was often diagnosed in Mexican American depressed patients who spoke English as a second language and antidepressants were not prescribed (8). Anxiety disorders are also underdiagnosed and sometimes confused with schizophrenia. Post-traumatic stress disorder (PTSD), once thought to be primarily a disorder of combat veterans, is common in rape victims and has been seen in inner city youth exposed to street violence (9). The presenting symptoms of PTSD, especially flashbacks and emotional blunting, may be easily confused with psychosis (9, 10). Obsessive-compulsive disorder is often not diagnosed among African Americans, yet there is not support for racial differences in prevalence (9). The ego dystonic thoughts or sometimes bizarre compulsive behavior could also be easily mislabeled as psychosis. Underrecognition of affective and anxiety disorders may mean that African Americans will not receive appropriate targeted treatment. Depression is often underrecognized in African Americans, either because of misdiagnosis, as noted above, or different treatment-seeking behavior. More than 50% of depressed African Americans may not receive treatment, partially because of racial differences in treatment-seeking behavior (11). Anxiety disorders may be underrecognized as well as misdiagnosed. African American patients with panic disorder often do not seek treatment (9). Those that do are often seen by general medical providers, who often manage panic disorder as a general medical condition. As noted above, obsessive-compulsive disorder and PTSD are often underrecognized as well as misdiagnosed. These disorders have an excellent prognosis when treated, but they have very poor outcomes, including the development of comorbid depression, loss of work, and social restriction, when not treated (9).

Differences in presenting symptoms may account for some of the misdiagnosing. African Americans tend to score differently on the Minnesota Multiphasic Personality Inventory showing more paranoid-type symptoms (11). More suspiciousness is seen in African Americans than in nonhispanic white populations, which some have interpreted as a "healthy paranoia" (2). Affective disorders are more likely to present with psychotic symptoms, mania may present with more irritable symptoms, and depression may present with suspiciousness (1, 13, 14). Moreover, therapists unfamiliar with African Americans may misrepresent cultural differences as psychopathology (15). Consequently, the risk for misdiagnosing schizophrenia is increased (15–17). Africans and perhaps African Americans are less likely to complain of guilt when they have an affective disorder (18). Presumably guilt is a more Western European concept because it assumes a more individualistic view of the world and focuses on individual responsibility. People from tribal cultures assume more communal responsibility; they are more likely to experience shame than guilt (18).

African Americans are treated differently in the mental health system. They are more likely to be hospitalized, to be involuntarily hospitalized, to sometimes receive seclusion and restraints, and to be discharged earlier (19–23). This differential treatment seems independent of degree of psychopathology (19). Not surprisingly, African Americans receive higher doses of antipsychotic medication, and more medication as requested by nurses (19, 25, 26). Careful trials have indicated that African Americans are more responsive to antipsychotic medication (5, 17, 27). Factors such as greater perceived dangerousness and social distance may play a part in the excessive use and dosing of antipsychotic medication (27). Many racial and ethnic minorities are reluctant to seek treatment in the standard mental health system. African Americans may be more likely to seek treatment or support from friends, minister, nonmental health professionals, or no treatment at all (15, 11). As noted above pharmacological treatment may not be received, but treatment could be delayed until symptoms become intolerable. The individual may be more symptomatic, which would result in a greater likelihood of hospitalization, involuntary treatment, and more medication (5, 17). However, as noted above, racial differences in degree of psychopathology are often not found. Reports in the literature on hypertension suggest that African Americans are more likely to be noncompliant with medication, perhaps because of socioeconomic factors, less access to care, and cultural beliefs about medication (28–30). (Research of this type is limited in psychiatry.) They tend to terminate or are dismissed from psychiatric treatment sooner than Caucasians (19, 25). Moreover, African Americans with psychiatric disorders may be less compliant with pharmacotherapy and African American males with schizophrenia are more likely to be on depot medication, which is medication given by injection lasting for weeks or months (31, 32).

African Americans generally perceive mental health services differently from Caucasians. They are more likely to fear treatment and to fear hospitalization (11). An active issue in the African American community is a strong awareness of the Tuskegee Study (33). This study, begun in the thirties, was federally sponsored and involved exclusively African American men in whom syphilis was diagnosed. The men were never provided with antibiotic therapy when it became available and were not informed. Partially because of a lack of knowledge or misinformation, African Americans perceive psychotropic medications as being experimental like the syphilis experiments or unacceptably mind altering. In addition, as already noted, distrust of standard mental health treatment is widespread, and treatment is often sought from non-mental health professionals. The distrust of the system is further exacerbated by social economic concerns; providers are often of a higher socioeconomic class than the patients or African Americans whose income is 60% that of Caucasians (34).

Recent studies suggest that Asians may require lower doses of tricyclic antidepressants and antipsychotics (35, 36). Presumably the cytochrome P450 microsomal enzyme system is altered so that these agents are metabolized more slowly than in Caucasians (37, 38). African Americans may also require lower doses of tricyclic antidepressants and antipsychotics although the evidence does not appear to be as strong (5, 17, 39). African Americans may respond more quickly to antipsychotic agents than Caucasians and may develop toxicity more easily on taking tricyclic antidepressants. Certainly there is no evidence that African Americans require more medication (5, 17). Yet, as noted above, African Americans may receive higher doses of antipsychotic medication than Caucasians. No literature exists for the dosing pattern of other medications. There are multiple reports of African Americans showing higher red blood cell-to-plasma levels of lithium than for Caucasians and Asians (40). The clinical significance of this finding is unknown but high red blood cell-to-plasma ratios are believed to be associated with lithium-neuroleptic toxicity (40). Recently, Strickland and associates (41) reported that the high red blood cell-to-plasma ratios seen in African Americans are associated with more reported lithium-related side effects. African Americans treated with antipsychotic medication may have a greater risk for tardive dyskinesia than Caucasians (42, 43). The reason for this increase is unknown at present. High doses of antipsychotics, long treatment duration, interruptions in treatment, and use of antipsychotic agents in those with affective disorders are all factors that increase the risk of tardive dyskinesia (44). As noted, these prescribing patterns are seen in African Americans, but whether there is an unspecified biological risk is presently unknown.

Newer treatments may be especially beneficial for African Americans. Newer antidepressants such as the specific serotonergic reuptake blockers may be beneficial for African Americans because of the reports of greater sensitivity or toxicity (45). Newer antipsychotics such as clozapine and risperidone may offer increased efficiency and reduced risk of abnormal involuntary movements including tardive dyskinesia (46, 47). Cost limits the availability of many of these agents for African Americans (48); African Americans often obtain treatment through the public sector but cost constraints necessarily cause these providers to be conservative about their treatment. The new antipsychotic clozapine has been shown to be more effective than standard antipsychotics and to be tolerated better by patients (46). However its cost, the most for any antipsychotic, greatly limited its availability. Although Medicaid sometimes made the agent available for the very poor, the working poor, not covered by Medicaid and frequently without health insurance, could not afford such medication. In fact the cost of having a family member with schizophrenia would approach the median income of African American families even without clozapine (5). Some of clozapine's high cost is caused by the increased risk it has of causing the potentially lethal complication of agranulocytosis (49). As a result, the Federal Drug Administration and Sandoz, Inc. (the agent's manufacturer) provide strict guidelines that require minimal leukocyte counts before treatment with the agent can be initiated or continued. However African Americans tend to show a relative and benign leukopenia (50). Thus individuals that might benefit from clozaril would not get it if the clozaril prescribing guidelines were strictly followed.

Conclusion

Racial and ethnic minorities are less likely than Caucasians to undergo some medical procedures and to be included in pharmaceutical trials (51, 52). We reviewed the psychiatric literature and found that most biological and pharmacological studies did not identify the ethnicity of the subjects (16). When race or ethnicity was identified, African Americans were in a decided minority. Consistent with reports from other disciplines, less than 6% of subjects were identified as African Americans. As noted, research is often viewed unfavorably in the African American community, but inclusion of racial and ethnic minorities in drug trials is extremely important to fully understand the impact of pharmacological agents on the general population. Moreover drug trials offer the additional benefit of making new treatments available at no cost. More must be done to educate the public, including administrators, providers, and patients, about ethnic diversity in pharmacotherapy. Racial and ethnic minorities in particular must be included in any educational program or marketing efforts that promote the use of psychotropic agents.

References

1. Adebimpe VR. Overview: White norms and psychiatric diagnosis of black patients. Am J Psychiatry 1981; 180: 279–285.
2. Jones BE, Gray BA. Problems in diagnosing schizophrenia and affective disorders among blacks. Hosp Community Psychiatry 1986; 37:61–65.
3. Neighbors HW. The distribution of psychiatric morbidity in black Americans: A review and suggestion for research. Community Ment Health J 1984; 20:169–181.
4. Kessler RC, McGonogle KA, Zhao S, et al. Lifetime and 12 month prevalence of DSM III-R psychiatric disorders in the United States. Arch Gen Psychiatry 1994; 51:8–19.
5. Lawson WB. Racial and ethnic factors in psychiatric research. Hosp Community Psychiatry 1986; 37:50–54.
6. Bell CC, Mehta H. The misdiagnosis of black patients with manic-depressive illness. J Natl Med Assoc 1980; 72:141–145.
7. Bell CC, Mehta H. Misdiagnosis of black patients with manic-depressive illness: second in a series. J Natl Med Assoc 1981; 73:101–107.
8. Lawson WB, Herrera JM, Costa J. The dexamethasone suppression test as an adjunct in diagnosing depression. J Assoc Acad Minor Phys 1992; 3:17–19.
9. Paradis CM, Hatch M, Friedman S. Anxiety disorders in African Americans: an update. J Natl Med Assoc 1994; 86:609–612.
10. Allen IM. Posttraumatic stress disorder among black Vietnam veterans. Hosp Community Psychiatry 1986; 37:55–61.
11. Sussman LK, Robins LN, Earls F. Treatment-seeking for depression by black and white Americans. Soc Sci Med 1987; 24:187–196.
12. Adebimpe VR, Gigardet J, Harris E. MMPI diagnosis of black psychiatric patients. Am J Psychiatry 1979; 135: 85–87.
13. Adebimpe VR, Hedlund Jl, Cho DW, et al. Symptomatology of depression in black and white patients. J Natl Med Assoc 1982; 74:185–190.
14. Adebimpe VR, Klein HE, Fried J. Hallucinations and delusions in black psychiatric patients. J Natl Med Assoc 1981; 73:517–520.
15. Adebimpe VR. Race, racism, and epidemological surveys. Hosp Community Psychiatry 1994; 45:27–31.
16. Lawson WB. Biological markers in neuropsychiatric disorders: racial and ethnic factors. In: Sorel E, editor. Family culture and psychobiology. New York: Leyas, 1990.
17. Strickland TK, Ranganath V, Lin K-M, et al. Psychopharmacologic considerations in the treatment of black American populations. Psychopharmacol Bull 1991; 27: 441–448.
18. German GA. Aspects of clinical psychiatry in Sub-Saharan Africa. Br J Psychiatry 1972; 121:461–479.
19. Flaherty JA, Meagher R. Measuring racial bias in inpatient treatment. Am J Psychiatry 1980; 137:679–682.
20. Lawson WB, Hepler N, Holladay J, et al. Race as a factor in inpatient and outpatient admissions and diagnosis. Hosp Community Psychiatry 1994; 45:72–74.
21. Herrera J. The effectiveness of cultural milieu on hospitalized Hispanic patients. Int J Psychosom 1987; 34:6–8.
22. Paul GI, Menditto AA. Effectiveness of inpatient treatment programs for mentally ill adults in public psychiatric facilities. Appl Prev Psychol 1992; 1:41–63.
23. Strakowski SM, Heather SL, Sax K, et al. The effects of race of diagnosis and disposition from a psychiatric emergency service. J Clin Psychiatry 1995; 56:101–107.
24. Lindsey KP, Paul GL, Mariotto MJ. Urban psychiatric commitments: disability and dangerous behavior of black and white recent admissions. Hosp Community Psychiatry 1989; 40:286–294.
25. Chung H, Mahler JC, Kakuna T. Racial differences in treatment of psychiatric inpatients. Psychiatry Services 1995; 46:586–591.
26. Strakowski SM, Shelton RC, Kolbrener ML. The effects of race and comorbidity on clinical diagnosis in patients and psychosis. J Clin Psychiatry 1993; 54:96–102.
27. Lawson WB, Yesavage JA, Werner RD. Race, violence, and psychopathology. J Clin Psychol 1984; 45:294–297.
28. Daniels DE, rene AA, Daniels VR. Race: an explanation of patient compliance—fact or fiction? J Natl Med Assoc 1994; 86:20–25.
29. Hildreth C, Saunders O. Hypertension in blacks: clinical overview. Cardiovasc Clin 1991; 21:85–96.
30. Saunders E. Hypertension in African Americans. Circulation 1991; 83:1465–1467.
31. Price N, Glazer W, Morgenstern H. Demographic predictors of the use of injectable versus oral antipsychotic medications in outpatients. Am J Psychiatry 1985; 142: 1491–1492.
32. Smith M, Lin K-M, Mendoza R. Nonbiological issues affecting psychopharmacological therapy: cultural considerations. In: Lin K-M, Poland RE. Nakasaki G, editors. Psychopharmacology and psychobiology of ethnicity. Washington DC: American Psychiatric Association, 1993: 37–58.
33. Roy B. The Tuskegee syphilis experiment: biotechnology and the administrative state. J Natl Med Assoc 1995; 87:56–67.
34. Lawson WB. The black family and chronic mental illness. Am J Social Psychiatry 1986; 6:57–61.
35. Lin K-M, Finder E. Neuroleptic dosage for Asians. Am J Psychiatry 1983; 140:490–491.
36. Yamashita I, Asano Y. Tricyclic antidepressants: thera-

peutic plasma level. Psychopharmacol Bull 1979; 5:40–41.

37. Bond WS. Therapy update: ethnicity and psychotropic drugs. Clin Pharmacol 1990; 10:467–470.
38. Lin K-M, Poland RE. Ethnicity, culture, and psychopharmacology. In: Bloom FE, Vand Kupler DJ, editors. Psychopharmacology: the fourth generation of progress. New York: Raven Press, 1995.
39. Silver B, Poland RE, Lin K-M. Ethnicity and the pharmacology of tricyclic antidepressants. In: Lin K-M, Poland RE, Wallasaki G, editors. Psychopharmacology and psychobiology of ethnicity. Washington DC: American Psychiatric Association, 1993: 61–89.
40. Strickland TL, Lawson WB, Lin K-M. Interethnic variation in response to lithium therapy among African-American and Asian-American populations. In: Lin K-M, Poland RE, Nakasaki G, editors. Psychopharmacology and psychobiology of ethnicity. Washington DC: American Psychiatric Association, 1993: 107–123.
41. Strickland TL, Lin K-M, Fu P, et al. Comparison of lithium ratio between African-American and Caucasian bipolar patients. Biol Psychiatry 1995; 37:325–330.
42. Glazer WM, Morgenstern H, Doucette J. Race and tardive dyskinesia among outpatients at a CMHC. Hosp Community Psychiatry 1994; 45:38–42.
43. Morgenstern H, Glazer WM. Identifying risk factors for tardive dyskinesia among chronic outpatients maintained on neuroleptic medications: results of the Yale Tardive Dyskinesia Study. Arch Gen Psychiatry 1993; 50:723–733.
44. Jeste DV, Wyatt RJ. Understanding and treating tardive dyskinesia. New York: Guilford, 1982.
45. Preskorn SH. Comparison of the tolerability of bupropion, fluoxetine, imipramine, nefazodone, paroxexine, sertraline, and verlafaxine. J Clin Psychiatry 1995; 56(suppl 6):12–21.
46. Kane J, Honifield G, Singer J, et al. Clozapine for the treatment of resistant schizophrenic: a double-blind comparison versus chlorpromazine. Arch Gen Psychiatry 1988; 45:789–796.
47. Marder SR, Meibach RC. Risperidone in the treatment of schizophrenia. Am J Psychiatry 1994; 151:825–835.
48. Revicki DA, Lou BR, Weschler JM, et al. Cost-effectiveness of clozapine for treatment-resistant schizophrenic patients. Hosp Community Psychiatry 1990; 41:850–854.
49. Alvir JM, Leiberman JA, Safferman AZ, et al. Clozapine-induced agranulocytosis: incidence and risk factors in the United States. J Med 1993; 329:162–167.
50. Caramikat E, Karayalein G, Aballi A, et al. Leucocyte count differences in healthy white and black children 1 to 5 years of age. J Pediatrics 1973; 86; 252–275.
51. Wenneker MB, Erstein AM. Racial inequalities in the use of procedures for patients with ischemic heart disease in Massachusetts. J Am Med Assoc 1989; 261:253–257.
52. Svenson CK. Representation of American blacks in clinical trials of new drugs. J Am Med Assoc 1989; 261:263–265.

Neuroleptic Dosing in Asian and Hispanic Outpatients with Schizophrenia

SIGFRIED RUIZ, M.D., PETER CHU, M.D., JOHN SRAMEK, PHARM.D., ECTERINA ROTAVU, M.D., AND
JOHN HERRERA, PH.D.

Abstract

Recent studies suggest that minorities may have higher plasma levels caused by ethnic variations in pharmacokinetics under standard neuroleptic treatment; they would be at increased risk for extrapyramidal side effects and more likely to be treatment noncompliant. This study examined the prescribing pattern of antipsychotic medication to minority schizophrenics in outpatient psychiatric programs developed to meet their treatment needs. A computer search was conducted of registered clients in a Hispanic and Asian outpatient clinic(s). Hispanic and Asian patients with schizophrenia were identified and for purposes of comparison, a third sample of consecutively drawn outpatients with a diagnosis of "schizophrenia" registered in the outpatient clinic proper were selected. All outpatient medical records were secured and neuroleptic dosing patterns extracted and quantified. Neuroleptic doses were converted to chlorpromazine equivalents and corrected for body weight to a standard of 68 kg. One-way analysis of variance procedures were used to compare both the actual and standardized neuroleptic doses across the three samples and revealed a significant main effect for both actual ($p < 0.05$) and standardized dose ($p < 0.05$). With regard to the former, secondary analysis yielded significant differences between the clinic sample and the Hispanic ($p < 0.05$) and Asian ($p < 0.05$) samples, which did not differ significantly from each other but which received significantly less drugs than the other patient population. These results were similar to the standardized dose comparison. The findings replicate previous inpatient reports and have implications for cross-cultural treatment programs.

SEVERAL retrospective surveys and reports (1–6), but not all (7), have suggested that Asian patients with schizophrenia need lower doses of neuroleptic drugs than Caucasians to achieve a clinical response. In addition, Asian patients may be more vulnerable to extrapyramidal reactions than Caucasians (8). A prospective study found that Asian patients with schizophrenia had higher mean haloperidol concentrations than Caucasians during a 2-week treatment period in which haloperidol was dosed by body weight, and also that Asians experienced a higher incidence of extrapyramidal symptoms during treatment despite benztropine routinely prescribed to both groups (9). Most of these reports, however, were based on inpatient dosing experiences and there is no information whether such cross-cultural differences found during acute or subacute treatment can actually translate to an outpatient setting. An early anecdotal report (3) listed various neuroleptic dosages prescribed to functionally psychotic Asian patients, but did not compare these dosages with a control group. This report is a retrospective survey of cross-cultural outpatient neuroleptic treatment which compares dosing in Asians with dosing in other ethnic groups from the same institution. A second objective of the study was to determine whether Hispanic patients with schizophrenia had neuroleptic dosage requirements different from either Asian or another control group. Although there is increasing

---

From the Mount Sinai School of Medicine and the Elmhurst Hospital Center. Address reprint requests to John Herrera, Ph.D., Associate Professor of Psychiatry, Mount Sinai Medical School and Associate Director of Psychiatry, Elmhurst Hospital Center, 79-01 Broadway, Elmhurst, NY 11373.

evidence of ethnic variations in psychiatric dosing requirements, there is a remarkable lack of data on Hispanics.

Methods

Site. This study was conducted in the Department of Psychiatry at Elmhurst Hospital Center (EHC, Elmhurst, NY), an academic affiliate of the Mount Sinai School of Medicine (MSSM, New York, NY). EHC is located in a large multiethnic community with a population that is approximately 40% Hispanic and 20% Asian. To serve this diverse clientele more adequately, the Department of Psychiatry (EHC) developed specialized clinical facilities to reflect this service community. These specialized facilities include an Hispanic inpatient unit and outpatient clinic and an Asian inpatient unit and outpatient clinic. Both are staffed by Hispanic and/or Asian psychiatrists, psychologists, social workers, nursing and support staff. This is the first of two studies (see Collazo et al. on p. 000 of this issue) to examine the prescribing of antipsychotic medication to patients with schizophrenia in these specialized facilities.

Subjects. A computer search was conducted of registered clients in the Hispanic and Asian outpatient clinics at EHC to identify all "schizophrenia" spectrum DSM III-R (Diagnostic and Statistical Manual of Mental Disorders. 3rd ed., rev.) diagnosed patients. In the Hispanic outpatient clinic, 27 patients with schizophrenia were identified (10 paranoid, 11 undifferentiated, and 6 schizoaffective). Demographic characteristics are listed in **table 1.** All of the subjects in the Hispanic sample had been born outside of the United States, the mean ± SD number of years in this country was 18 ± 8.44 (range, 1–31). In the Asian outpatient clinic, 48 schizophrenia spectrum outpatients were identified (28 paranoid, 11 unidentified, and 9 schizoaffective). Three Asian patients were born in the United States. Of those born elsewhere ($n = 45$), the mean number of years ± SD in this country was 13.8 ± 8.85 (range, 1–42).

TABLE 1
Summary of Biodemographic Variables and Psychiatric History

	Hispanics	Asians	General	χ^2
Male	11	21	26	
Female	16	27	38	0.17
United States	0	3	30	
Other	28	45	30	35.92**
	Hispanics	Asians	General	*F*
Mean years in residence	18.00	13.82	13.82	1.94
Mean age in years	42.82	39.08	38.17	1.65
Mean years since onset	13.00	10.58	11.30	1.65
Mean psychiatric hospitalizations	3.09	1.68	3.00	5.64*

* $p < 0.01$.
** $p < 0.001$.

For the purpose of comparison, a third sample of "schizophrenia" spectrum DSM III-R diagnosed outpatients registered in the EHC outpatient clinic were selected. Subjects in the "general" sample were drawn consecutively beginning with the most recent admission after a designated date, excluding Asian and Hispanic patients with schizophrenia. Initially, 74 subjects with schizophrenia were selected to match the number of previously identified Hispanic and Asian patients with schizophrenia ($n = 26$ and 48, respectively); however, only 64 medical records were located or found to have sufficient data to be included in the study. These outpatients included 1 disorganized, 36 paranoid, 4 schizophreniform, 1 residual, 17 undifferentiated, and 5 schizoaffective patients. Thirty patients in the general sample were born in the United States; and, of those born elsewhere ($n = 34$), the mean ± SD number of years in this country was 13.82 ± 8.03 (range, 1–37).

Procedures. All outpatient medical records were solicited for the purpose of extraction and quantification of selected biodemographic, psychiatric history, and neuroleptic dosing patterns. These variables included the subjects' sex, height, weight, age, birthplace, years in residence (if applicable), confirmation of diagnosis (treating psychiatrist's most current progress notes), previous psychiatric hospitalizations, the number of years since the onset of psychosis, type of neuroleptic used, the current dose, treatment with other psychiatric medications (i.e., antidepressants), treatment with antiparkinson medications, and the presence of a documented movement disorder. All patients had been stabilized on their neuroleptic dose for at least 60 days and no patients were included who were receiving either risperidone or clozapine. For purpose of comparison, the neuroleptic doses were converted to chlorpromazine equivalents (10) and also corrected for body weight to a standard of 68 kg. Both the actual and the 68-kg standardized dose were used in subsequent analyses.

Results

χ^2 comparisons failed to reveal significant gender differences among the three outpatient schizophrenic samples (see Table 1). Similar comparisons did reveal large differences ($\chi^2 = 35.92$, $p < 0.0001$) between the number of Hispanic, Asian, and general sample patients born in this country or elsewhere. Although approximately half of the general sample had been born in the United States, most Hispanic and Asian patients were born elsewhere. However, one-way analysis of variance (ANOVA) procedures failed to yield significant differences among the three samples in the number of years in residence in the United States (only those general sample patients born outside of the United States were included in this analysis, $n = 33$). Similar procedures (ANOVA) failed to yield significant age differences among the three outpatient schizophrenic groups nor were significant differences found in the number of years since the onset of psychosis. Significant differences were found among the three samples in the number of previous psychiatric hospitalizations ($F = 5.64$, $p < 0.004$) and subsequent within-sample comparisons revealed differences between the Asian and Hispanic samples ($p < 0.009$) and the Asian and General sample ($p < 0.004$). Previous reports have noted that Asians tend to shun psychiatric treatment until late in the course of their disorder (9).

ANOVA procedures were used to compare both the actual and standardized neuroleptic doses across the three samples (Hispanic, Asian, and general). As seen in **Table 2**, these analyses revealed a significant main effect for both actual ($F = 5.68$, $p < 0.004$) and standardized dose ($F = 4.79$, $p < 0.009$). With regard to the former, secondary analysis yielded significant differences between the general sample in comparison with the Hispanic ($p < 0.029$) and Asian ($p < 0.002$) samples, which did not differ significantly from each other. These results were similar to the standardized dose comparisons (see Table 2). Significant differences between the general sample and the two ethnic minority samples were found ($p < 0.029$ for Hispanics and $p < 0.008$ for Asians), although neither differed significantly from each other. Examination of the direction of mean differences for both medication dosing variables revealed that the general sample was prescribed significantly larger doses of antipsychotic medication than either of the two ethnic minority samples. χ^2 comparisons failed to reveal significant differences among the three samples in the prescription of other psychiatric medications nor in the presence of a documented movement disorder. However significant differences were found in the prescription of antiparkinson medication ($\chi^2 = 22.69$, $p < 0.0001$). As seen in Table 2, Asian patients were prescribed antiparkinson medication significantly more often than subjects in either of the other two samples. To further understand these dosing differences, neuroleptics were differentiated on the basis of potency (high vs low) and χ^2 tests revealed significant differences among the three samples ($p < 0.035$). As seen in Table 2, both ethnic minority groups were prescribed more low-potency medications than the general sample.

TABLE 2

Summary of Mean Actual and Standardized CPZ[a] *Dose, Psychiatric Medication, Movement Disorder, and Antiparkinson Medication*

	Hispanics	Asians	General	F
Mean actual CPZ dose	223.67	194.89	347.33	5.68**
Mean standard CPZ dose	211.53	212.26	339.60	4.79*

	Hispanics		Asians		General		
	Yes	No	Yes	No	Yes	No	χ^2
Other psychiatric medication	8	18	9	39	16	48	1.76
Movement disorder	1	25	2	46	1	62	0.74
Antiparkinson medication	11	15	42	6	33	30	22.69***

	Hispanics		Asians		General		
	Hi	Low	Hi	Low	Hi	Low	χ^2
Other psychiatric medication	20	6	37	9	56	3	6.72*

[a] CPZ, chlorpromazine.
* $p < 0.01$.
** $p < 0.001$.
*** $p < 0.0001$.

Discussion

Our retrospective survey of neuroleptic dosing practices with outpatients who have schizophrenia confirmed previous reports (1–8) that Asians may require lower dosage. However, an unexpected finding was that Hispanic patients with schizophrenia also received significantly lower dosages than the general sample at this institution. Hispanics have not been studied systematically regarding neuroleptic dosing requirements and sensitivity to extrapyramidal symptoms. In a study of tardive dyskinesia prevalence by Sramek et al. (11), inpatient Hispanic psychiatric patients having various diagnoses (including schizophrenia, mania, and depression) did not receive different neuroleptic dosage from Caucasian or African American patients. The overall similar use of low-potency neuroleptic medications between Hispanics and Asians in this study contrasts with a lower use of these drugs in the general sample and suggests that Hispanics may also be more sensitive to extrapyramidal symptoms than others. Nonetheless, similar criticisms can be levied at the retrospective methods used in this report as well as other retrospective surveys published previously. These include lack of patients from different ethnic groups matched for severity of illness, treatment time period, prior history (in particular prior use of neuroleptics), and differing prescribing practices by clinicians. Such surveys must then be confirmed by carefully conducted prospective studies which control as many of these factors as possible and which ideally randomize patients to fixed dosages of neuroleptics and prospectively rate patients for improvement and appearance of extrapyramidal symptoms. Pharmacokinetic assessments would also be highly desirable in such studies to investigate potential differences in drug metabolism among ethnic groups. Finally, these findings support recent cross-cultural research that advocates tailoring the therapeutic environment to be more “culturally responsive” to ethnic minorities.

References

1. Okuma T. Differential sensitivity to the effects of psychotropic drugs: psychotics vs normals; Asians vs Western populations. Folia Psychiatr Neurol Jpn 1981; 35:79–82.
2. Olatawara M. The effects of psychotropic drugs in different populations. Bull WHPP 1978; 56:519–523.
3. Yamamoto J, Fung D, Lo S, et al. Psychopharmacology for Asian Americans and Pacific Islanders. Psychopharmacol Bull 1979; 5:29–31.
4. Lin K, Finder E. Neuroleptic dosage for Asians. Am J Psychiatry 1983; 140:490–491.
5. Rosenblatt R, Tang S. Do Oriental psychiatric patients receive different dosages of psychotropic medication when compared with Occidentals? Can J Psychiatry 1987; 32:270–274.
6. Potkin S, Shen Y, Pardes H, et al. Haloperidol concentrations elevated in Chinese patients. Psychiatry 1984; 12: 167–172.
7. Sramek J, Sayles M, Simpson G. Neuroleptic dosage for Asians: a failure to replicate. Am J Psychiatry 1986; 143:535–536.
8. Binder R, Levy R. Extrapyramidal reactions in Asians. Am J Psychiatry 1981; 38:1243–1244.
9. Lin K, Poland R, Nuccio I, et al. A longitudinal assessment of haloperidol doses and serum concentrations in Asian and Caucasian schizophrenic patients. Am J Psychiatry 1989; 146:1307–1311.
10. Davis J. Dose equivalence of the antipsychotic drugs. J Psychiatr Res 1974; 11:65–69.
11. Sramek J, Roy S, Ahrens T, et al. Prevalence of tardive dyskinesia among three ethnic groups of chronic psychiatric patients. Hosp Community Psychiatry 1991; 42:590–592.

Neuroleptic Dosing in Hispanic and Asian Inpatients with Schizophrenia

Yasmine Collazo, M.D., Raymond Tam, M.D., John Sramek, Pharm.D., and John Herrera, Ph.D.

Abstract

This study is the second in a series examining the prescribing of antipsychotic medication to patients with schizophrenia in cross-cultural clinical programs. A computer search identified all patients with the diagnosis of "schizophrenia" treated during a 1-year period in an inpatient Hispanic and Asian psychiatric unit(s); second computer search identified a matched (admission date) sample of Anglo patients from the general inpatient psychiatry services. The medication variables included type of neuroleptic drug used, the maximum dose, the stabilized dose (i.e., neuroleptic dose at discharge) and the dose associated with first report of extrapyramidal symptoms. Neuroleptic doses were converted to chlorpromazine (CPZ) equivalents and corrected for body weight to a standard of 68 kg. One-way analysis of variance procedures were used to compare both actual and standardized neuroleptic CPZ across the three samples; these statistical comparisons were completed for both maximum and stabilized dose(s). The analysis with maximum dose revealed a significant main effect for both actual ($p < 0.05$) and standardized CPZ ($p < 0.05$). Similar results were also found for stabilized dose with both actual ($p < 0.05$) and standardized CPZ ($p < 0.05$). Examination of the direction of mean differences for both medication dosing variables using both CPZ comparisons revealed that the patients in general sample received significantly larger doses of antipsychotic medication than either Asian or Hispanic patients.

THIS REPORT IS A FOLLOW-UP survey to our outpatient neuroleptic dosing study in Asian and Hispanic patients with schizophrenia (see Ruiz et al., p. 306 of this issue). Results from the previous outpatient study confirmed earlier reports (2–9) that Asians may require lower dosage than other ethnic groups. A surprising finding, however, was that Hispanic patients also receive lower doses of neuroleptic drugs than other patients. The present study was conducted on the corresponding cross-cultural inpatient units at our institution for comparison with these outpatient findings.

Methods

Site. This study was conducted in the Department of Psychiatry at Elmhurst Hospital Center (EHC, Elmhurst, NY), an academic affiliate of Mount Sinai School of Medicine (MSMS, New York, NY). EHC is located in a large multiethnic community in the borough of Queens, NY and to serve this diverse clientele more adequately, the Department of Psychiatry developed inpatient and outpatient Hispanic and Asian programs, both staffed by Hispanic and/or Asian psychiatrists, psychologists, social workers, nursing, and support staff.

Subjects. A computer search of EEHC hospital records was conducted to identify all "schizophrenia" spectrum DSM III-R (Diagnostic and Statistical Manual of Mental Disorders, 3rd ed., rev.) diagnoses patients (discharge diagnosis) treated during a 1-year period in inpatient Hispanic and Asian psychiatric units. Twenty-seven

---

From the Mount Sinai School of Medicine and the Elmhurst Hospital Center. Address reprint requests to John Herrera, Ph.D., Associate Professor of Psychiatry, Mount Sinai School of Medicine and Associate Director of Psychiatry, Elmhurst Hospital Center, 79-01 Broadway, Elmhurst, NY 11373.

Hispanic patients with schizophrenia were identified in the former and 14 Asian patients with schizophrenia were identified in the latter. For the purpose of comparison, a second computer search was completed to identify a matched (admission date) sample of 41 Anglo patients from the EHC inpatient psychiatry services. Examination of medical records (see "Procedures") revealed that three Anglo subjects had been misidentified (two Indian and one African American), two charts were found to have insufficient data and one chart was not found; therefore, these patients were excluded from the data analysis. Although the Asian sample included 10 Chinese and 4 Korean subjects, only the Chinese subjects were included in this study because the Korean subjects will be included in a subsequent (ongoing) report. This study thus included 27 Hispanic, 10 Asian (Chinese), and 35 Anglo subjects. The inpatient Hispanic schizophrenic sample included 1 disorganized, 14 paranoid, 2 schizophreniform, 1 residual, 5 schizoaffective, and 4 undifferentiated types. Demographic characteristics are listed in **Table 1.** With the exception of three subjects, all Hispanic patients were born outside of the United States, and the mean ± SD number of years in this country was 13.65 ± 11.2 (range, 1–41). The inpatient Asian schizophrenic sample included 3 paranoid, 1 schizophreniform, 3 residual, 2 schizoaffective, and 1 disorganized type. All Asian patients were born outside of the United States; the mean ± SD number of years in this country was 8.60 ± 7.0 (range, 1–24). The inpatient Anglo schizophrenic sample included 19 paranoid, 7 residual, 5 schizoaffective, and 4 undifferentiated types. Twenty-four Anglo patients were born in the United States and, of the 11 Anglo patients born elsewhere, the mean ± SD number of years in this country was 18.01 ± 12.4 (range, 4–38).

TABLE 1
Summary of Biodemographic Variables and Psychiatric History

	Hispanics	Asians	General	χ^2
Male	4	0	18	
Female	23	10	16	15.44**
United States	3	0	24	
Other	24	10	10	28.43**
	Hispanics	**Asians**	**Anglo**	***F***
Mean years in residence	13.65	8.60	18.01	
Mean age in years	36.07	38.40	45.45	4.04*
Mean years since onset	7.28	9.3	17.05	8.28**
Mean psychiatric hospitalizations	3.50	2.20	4.55	

* $p < 0.01$.
** $p < 0.001$.

Procedures. All medical records depicting the inpatient experience of the three samples were elicited for the purpose of extraction and quantification of selected biodemographic, psychiatric history, and neuroleptic dosing patterns. These variables included the subject's sex, age, birthplace, years in residence (if applicable), previous psychiatric hospitalizations, the number of years since the onset of psychosis, height, weight, confirmation of diagnosis (discharge summary), secondary psychiatric diagnosis (i.e., substance abuse), type of neuroleptic drug(s) used, the maximum dose, the stabilized dose (i.e., neuroleptic dose at discharge), the dose associated with first report of extrapyramidal symptoms, treatment with other psychiatric medications (i.e., antidepressants), the presence and type of documented movement disorder, treatment with antiparkinson medications, and whether this medication was started prophylactically or only after extrapyramidal symptoms were detected. For purpose of comparison, the neuroleptic doses were converted to chlorpromazine (CPZ) equivalents (7) and also corrected for body weight to a standard of 68 kg. Both the actual and the standardized doses were used in subsequent analyses. No patients who were receiving either risperidone or clozapine were included in this analysis.

Results

χ^2 comparisons revealed significant gender differences among the three samples ($p < 0.01$); as seen in Table 1 both ethnic minority samples were predominantly female. One-way analysis of variance (ANOVA) procedures revealed significant age differences ($F = 4.05$, $p < 0.021$); however, secondary comparisons yielded significant differences ($p < 0.007$) only between the Hispanic and Anglo samples (see Table 1). χ^2 comparisons also revealed large differences ($\chi^2 = 28.44$, $p < 0.0001$) between the number of Hispanic, Asian, and Anglo patients born in this country or elsewhere. Although approximately two thirds of the Anglo sample had been born in the United States, most Hispanic and Asian patients were born elsewhere. However, ANOVA procedures failed to yield significant differences among the three samples in the number of years in residence in the United States (only those Anglo patients born outside of the United States were included in this analysis, $n = 11$). Similar ANOVA calculations failed to yield significant differences in the number of previous psychiatric hospitalizations among the three samples; however, significant differences were revealed ($F = 8.28, p < 0.006$) in years since the onset of psychosis. Subsequent within-sample comparisons yielded significant differences between the Anglo sample in comparison with the Hispanic ($p < 0.026$) and Asian ($p < 0.002$) samples, which did not differ significantly from each other.

ANOVA procedures were used to compare both actual and standardized neuroleptic CPZ doses across the three samples (Hispanic, Asian, and Anglo). These statistical comparisons were completed for both maximum and stabilized doses. As seen in **Table 2**, the analysis with maximum dose revealed a significant main effect for both actual ($F = 6.74, p < 0.002$) and standardized CPZ ($F = 3.18, p < 0.047$). Similar results were also found for stabilized doses with both actual ($F = 5.92, p = 0.004$) and 68-kg standardized CPZ ($F = 3.26, p = 0.044$). Regarding the maximum dose and actual CPZ dose, secondary analysis yielded significant differences between the Anglo sample in comparison with the Hispanic ($p < 0.009$) and Asian ($p < 0.002$) samples, which did not differ significantly from each other. These results were similar with maximum dose and 68-kg standardized CPZ comparisons (see Table 2); significant differences between the Anglo sample and the two ethnic minority samples were found (Hispanic vs Anglo, ($p < 0.036$; Asian vs Anglo, $p < 0.056$), although neither differed significantly from each other. Secondary analysis with actual CPZ and stabilized doses also yielded significant differences between the Anglo sample in comparison with the Hispanic ($p < 0.017$) and Asian ($p < 0.003$) samples, which did not differ significantly from each other. These results were similar to 68-kg standardized CPZ and stabilized dose comparisons (see Table 2: Hispanic vs Anglo, $p < 0.057$; Asian vs Anglo $p < 0.033$). Examination of the direction of mean differences for both medication dosing variables (i.e., maximum and stabilized dose) using both CPZ comparisons (i.e., actual and standardized) revealed that significantly larger doses of antipsychotic medication were prescribed for patients in the Anglo sample than either of the two ethnic minority samples.

TABLE 2

Summary of Mean Actual and Standardized CPZ Dose, Psychiatric Medication, Movement Disorder, and Antiparkinson Medication

	Hispanics	Asians	Anglo	F
Mean maximum actual CPZ dose	527.59	392.50	726.00	6.74**
Mean maximum standard CPZ dose	353.44	487.19	712.23	3.18*
Mean stabilized actual CPZ dose	396.56	257.50	599.14	5.92**
Mean stabilized standard CPZ dose	411.78	37.89	593.00	3.26**

	Hispanics		Asians		Anglo		
	Yes	No	Yes	No	Yes	No	χ^2
Other psychiatric medication	21	06	5	5	17	18	5.86*
Movement disorder	12	15	0	10	3	32	14.94**
Antiparkinson medication	24	3	8	2	32	2	

* $p < 0.01$.
** $p < 0.001$.

χ^2 comparisons (see Table 2) revealed significant differences among the three samples in the prescription of other psychiatric medications ($\chi^2 = 5.86, p < 0.053$); however, differentiating these medications by type (antidepressant, anxiolytic, antiseizure, lithium, hypnotic) did not yield significant differences (because of the number of subsequent cells and the small sample sizes). χ^2 examination of the use of antiparkinson medications failed to reveal significant differences among the three samples, and there were no differences in the use of these medications prophylactically or only after extrapyramidal side effects had been detected. With few exceptions, subjects in all samples had been treated prophylactically with antiparkinson medication. Highly significant differences were found in the diagnosis of movement disorders ($\chi^2 = 14.94, p < 0.0006$); and, as seen in Table 2, these resulted from the presence of movement disorders in approximately half of the Hispanic sample. Subsequent examination of the neuroleptic dosage (actual and standardized) at discharge (stabilized dose) of those patients with documented movement disorder (i.e., EPS, tardive dyskinesia) did not reveal significant ethnic differences.

Discussion

Although inpatient subjects with schizophrenia are treated by different clinicians from those treating outpatients at our institution, the inpatient survey confirmed the findings that both Asians and Hispanics receive significantly lower doses than other patients. Unfortunately, significant biodemographic differences in age, sex distribution, and duration of illness between ethnic and Anglo patients limit confidence in the results, but do suggest as an independent indicator that both ethnic groups deserve further study. Carefully controlled and conducted prospective studies, particularly focused to investigate dosing in Hispanics, should be conducted to confirm these preliminary findings. Such studies need to evaluate the methods for assessing treatment outcomes, such as severity of illness and improvement which may be influenced by ethnic biases in disease evaluation. One of the greatest barriers to further transcultural psychopharmacology research is that the minimum effective dose of neuroleptic drugs remains unknown, despite the availability of these drugs for nearly four decades (8, 9). In an early elegant study by Wode-Helgodt et al. (10), improvement was positively related to a dose range of chlorpromazine which encompasses the CPZ equivalent dose spectrum found in this spectrum. Thus, an alternative hypothesis is that Anglos may also respond to the lower neuroleptic doses found effective with Asians and Hispanics. Use of fixed doses of an antipsychotic agent which has a well demonstrated dose-response relationship is crucial for future cross-cultural psychopharmacology research.

References

1. Okuma T. Differential sensitivity to the effects of psychotropic drugs: Psychotics vs normals; Asians vs Western populations. Folia Psychiatr Neurol Jpn 1981; 35:79–82.
2. Olatawara M. The effects of psychotropic drugs in different populations. Bull WHPP 1978; 56:519–523.
3. Yamamoto J, Fung D, Lo S, et al. Psychopharmacology for Asian Americans and Pacific Islanders. Psychopharmacol Bull 1979; 15:29–31.
4. Lin K, Finder E. Neuroleptic dosage for Asians. Am J Psychiatry 1983; 140:490–491.
5. Rosenblatt R, Tang S. Do Oriental psychiatric patients receive different dosages of psychotropic medication when compared with Occidentals? Can J Psychiatry 1987; 32:270–274.
6. Potkin S, Shen Y, Pardes H, et al. Haloperidol concentrations elevated in Chinese patients. Psychiatry 1984; 12: 167–172.
7. Davis J. Dose equivalence of the antipsychotic drugs. J Psychiatr Res 1974; 11:65–69.
8. Janicak P, Javaid R, Sharma A, et al. Random assignment to three haloperidol plasma levels for acute psychosis. Biol Psychiatry 1994; 35:666.
9. Stone C, Garver D, Griffith J, et al. Further evidence of a dose-response threshold for haloperidol in psychosis. Am J Psychiatry 1995; 152:1210–1212.
10. Wode-Helgodt B, Borg S, Fyro B, Sedvall G. Clinical effects and drug concentrations in plasma and cerebrospinal fluid in psychotic patients treated with fixed doses of chlorpromazine. Acta Psychiatr Scand 1978; 58:149–173.

Ethnicity and Movement Disorders

John H. Eastham, Pharm.D., Jonathan P. Lacro, Pharm.D., and Dilip V. Jeste, M.D.

Abstract

This article is a review of studies involving ethnicity and spontaneous as well as iatrogenic movement disorder. We have focused on Parkinson's disease and tardive dyskinesia. Some early studies reported that African Americans had lower rates of Parkinson's disease than Causasians, whereas more recent surveys have not found such a difference. Several studies suggested that African Americans have a higher risk of developing tardive dyskinesia than Caucasians, even when differences in neuroleptic drug use are accounted for. Asians seem to have a lower or equal risk of developing tardive dyskinesia as compared with Caucasians. We discuss the possible contribution of ethnicity to the etiology of movement disorders and the implications thereof.

Several epidemiologic studies have examined the role of ethnicity in movement disorders, especially Parkinson's disease (PD) and tardive dyskinesia (TD). Of the spontaneous movement disorders, PD is one of the most important conditions. Neuroleptic-induced TD is the only iatrogenic movement disorder in which the dyskinesia can persist even after withdrawal of neuroleptic medications. (1). Here, we review the relevant literature on these disorders.

Parkinson's Disease

The epidemiology of idiopathic PD may be of special interest to psychopharmacologists, because some protective and risk factors for this disease could be related to those for the neurologic-induced parkinsonism. Several epidemiologic studies in the 1970s indicated that PD was much less prevalent among African Americans than among Caucasians (2–5). In addition, community-based epidemiologic studies have generally shown that there is a higher prevalence of PD in Europe and North America than in Africa and Asia (6–9). These observations have led to the hypothesis that higher melanin pigmentation may play a protective role because PD is associated with a loss of melanin pigmentation in the substantia nigra. One must be cautious in interpreting and comparing the findings from these studies, however, because many factors influence the results. Differences in population study methodology, diagnostic criteria, and percentage of the population not participating may all lead to different findings.

In one of the earlier investigations of PD epidemiology examining the association of ethnicity and PD, Kessler (2) evaluated records for hospitalizations of patients with a diagnosis of PD, paralysis agitans, or parkinsonism from 1965 through 1967. Blinded interviews were conducted (if possible) with patients and matched controls, and medical records were reviewed to validate the presence or absence of PD in different groups. The overall prevalence rates of PD in men were 4.58/100,000 and 28.6/100,000 for African Americans and Caucasians, respectively. Rates were similar in women (4.28/100,000 for African Americans and 22.88/100,000 for Caucasian women). Simi-

---

This work was supported, in part, by National Institute for Mental Health grant 5 RO1 MH45131, the Department of Veterans Affairs, and by an educational grant from Eli Lilly and Companies. From University of California, San Diego and Geriatric Psychiatry Clinical Research Center, Veterans Affairs Medical Center, San Diego, California. Address reprint requests to Dilip V. Jeste, M.D., Professor of Psychiatry, University of California, San Diego and Director, Geriatric Psychiatry Clinical Research Center (116A1), Veterans Affairs Medical Center, 3350 La Jolla Village Drive, San Diego, CA 92122.

lar results were obtained from a survey conducted at a public hospital with a disproportionate (75%) African American population. One might argue that extrapolating these observations to the general population would be invalid, because the survey was based on hospitalized populations, and African Americans probably had inadequate access to hospital care (4). Kessler conducted a physician referral-based PD survey in Baltimore which also reported lower prevalence rates in African Americans (3). More recent epidemiologic studies which used better surveying techniques, however, argue against a link between ethnicity and PD. One such investigation of PD prevalence was conducted by Schoenberg et al. (10) in the biracial community of Copiah County, Mississippi. Copiah County is unique in that it has roughly equal numbers of African Americans and Caucasians (49.1% and 50.1%, respectively). Ninety-seven percent of the county's households participated in the interview, and 85% of the persons with responses suggestive of PD or other disorders of interest allowed a neurologic examination for confirmation. Using the least stringent diagnostic criteria (possible PD) the prevalence rate was similar in both groups (338/1000,000 and 353/100,00 for African Americans and Caucasians, respectively). Caucasians, however, had a higher age-adjusted prevalence rate if the more stringent criteria (definite PD) were used (196/100,000 and 280/100,00 for African Americans and Caucasians, respectively). Although this difference in definite PD may seem dramatic, it is not nearly as large a difference as suggested by Kessler (2, 3). One important finding of this countywide survey was the amount of undiagnosed PD found, especially in the African American population. More than one half (58%) of the African Americans and approximately one third (32%) of the Caucasians were found to have definite or possible PD which had previously been undiagnosed.

Further evidence that there is no link between PD incidence and ethnicity comes from Canada. Because Canada provides universal health care, equal access to health care and specialists can be consistent for all ethnic groups, a more direct comparison of PD incidence between racial groups can be made by health care utilization surveys (11). Although the indigenous population of Saskatchewan is well known for being resistant to multiple sclerosis, there was no difference in the incidence of PD in the indigenous versus Caucasians Saskatchewan population (rates not reported) (11). To compare the prevalence of PD in the African American populations living under different environmental conditions, Schoenberg et al. (12) conducted a door-to-door survey in Copiah County, Mississippi, and Igbo-Ora, Nigeria. Persons in a household found to have signs, symptoms, or findings suggestive of neurologic diseases were examined by a neurologist for the assessment of possible PD, or were asked to complete a more extensive questionnaire if they resumed the examination. Cooperation was excellent in both communities. After age adjustment, the prevalence rates were dramatically different. The overall prevalence rates were 67/100,000 and 341/1000,000 persons for Igbo-Ora and Copiah country population, respectively. When these results are viewed in light of the similar prevalence of PD in the African American and Caucasians populations of Copiah County (10), one is led to conclude that the etiology of PD is partly caused by environmental agents and that ethnicity is not a strong protective or predisposing factor.

Summary. Epidemiologic surveys that are based on health care service utilization are relatively inexpensive and easy to conduct. They are, however, likely to miss early cases and those not receiving medical care for the disease in question. With the exception of differences in population cooperation, diagnostic criteria, and diagnostic skills, door-to-door surveys are much more likely to provide the most accurate information of disease prevalence in the population surveyed (13). The work by Schoenberg et al. (10), which used a survey of almost all of Copiah County population, reported nearly equal rates for African Americans and Caucasians, and provided evidence of substantial underdiagnosis of PD, especially in the African American population (58% of African Americans and 32% of Caucasians previously undiagnosed). These are both important findings because they case doubt on the prevalence rates derived from hospital surveys. The lower incidence of PD in African Americans reported in earlier studies was probably caused by the fact that these studies used hospital and physician-based survey methods; these studies probably underestimated the true prevalence of PD in the African American population because of considerable underdiagnosis. In Saskatchewan, where access to health care was without apparent economic biases, there was no difference in the incidence of PD between the indigenous and Caucasians populations (11). The comparison between Nigeria and Copiah County revealed that Africans living in different environments had different prevalence rates and supported the hypothesis that PD is largely linked to environmental exposure (12). It is thus important to consider methodologic aspects of the studies as well as the possible influence of environmental factors on what may appear to be ethnic differences in the prevalence of a disease.

Huntington's Disease

Huntington's Disease (HD) is an inherited neuropsychiatric disorder which includes presentation of dyskinesia, monophasic dementia, and mood disorders. Several studies have reported that HD is rare among Africans and African Americans (14–16). A statewide survey in Maryland, however, revealed nearly equal prevalence of HD in African Americans and Caucasians (6.37/100,000 and 4.9/100,000, respectively) (17). African Americans had an earlier age of HD onset and suffered from more severe bradykinesia and eye movement abnormalities than Caucasians. The psychiatric disorders which frequently occur in HD were, however, found to be relatively rare in the African Americans. The discrepancy of the results between the Maryland survey and previous studies is most likely caused by the earlier studies having used less sophisticated methodology and not having reported clinical features.

Neuroleptic-Induced Tardive Dyskinesia

TD is a movement disorder lasting at least a few weeks, which is characterized by involuntary choreiform, athetoid, or rhythmic movements of the tongue, jaw, or extremities developing in association with the use of neuroleptic medication (18). In spite of countless studies of TD, pathophysiology of this drug-induced disease remains poorly understood (1). The role of ethnicity as a risk factor for TD is of interest because increasing our understanding of the risk factors for this movement disorder could potentially help reduce its incidence. One must be particularly cautious with epidemiologic studies involving TD. Population characteristics such as age, gender, and previous neuroleptic exposure should be taken into account as they are all reported to influence the risk of development of TD. Another important factor is, if the patients were receiving neuroleptic medications at the time of their examination, TD might be masked, which would result in underestimating its prevalence.

A review of 76 TD prevalence studies has been published by Yassa and Jeste (19). As part of a multifaceted review, the authors examined the TD prevalence studies published through 1989, and grouped them according to the continent of the study population. The overall means of TD prevalence from the individual continents were as follows: North America 27.6%, Europe 21.5%, Africa 25.5%, and Asia 16.6%. The authors explained the relatively low prevalence rate found in the Asian studies as possibly being related to the use of lower neuroleptic doses. An obvious limitation of using aggregated study results is that it does not take into account differences in the population sampled, definitions of TD, neuroleptic doses, or durations of neuroleptic exposure, etc. (19).

Studies in Asians. Hayashi et al. (20) examined the use of neuroleptic medications in Japan and incidence of TD in 73 neuroleptic-naive elderly patients. The study group consisted of 32 men and 41 women who had a mean ± SD age of 76.4 ± 6.8 (range, 65–91) years. Diagnoses included multi-infarct dementia (52%), delusional disorder (19%), Alzheimer's disease (14%), mood disorders (11%), alcohol dementia (4%), and other conditions (3%). These patients were compared with 74 demented nursing home patients (age 82.4 ± 6.8 years) who were neuroleptic-naive. Every 6 months the Abnormal Involuntary Movement Scale (AIMS) (21) was administered to these patients to assess abnormal movement. The mean ± SD study duration was 20.1 ± 40.2 (range, 1–244) months. Forty-four percent of the treatment group met TD diagnostic criteria, whereas only 14% of the control group developed spontaneous dyskinesias ($p < 0.0001$). The mean time for TD to develop was 35 months. Risk factors for developing TD were longer periods of hospitalization, and longer duration of neuroleptic treatment ($p < 0.01$). The TD rate of 44% reported in this study was similar to that recently reported by Jeste et al. (22) in a study of older patients in the United States.

Ogita et al. (23) reviewed the prevalence of TD in two hospitals in France and Japan. The French study group consisted of 131 subjects (75 male, 53 female), and the Japanese group included 123 subjects (70 male, 53 female). The criteria for TD diagnosis in this study were not mentioned, however, the examinations were performed by the same clinician. Overall prevalence rates of TD at the two hospitals were almost identical (18.3% and 17.9% in France and Japan, respectively). The French hospital had a higher prevalence rate of TD among its female patients (10.7% males and 28.6% females) than the Japanese hospital (17.1% males and 18.9% females). The French study group had more pronounced and less intermittent symptomatic dyskinesias than the Japanese group. Tongue movement, the most common form of orofacial dyskinesia, was also observed in all of the French cases and in only one half of the TD cases found in the Japanese group. These similar TD rates in Asians and Caucasians are in contrast to two previous studies which reported low rates in Asians. Ko et al. (24) examined 866 psychiatric inpatients from Shanghai, China, for the prevalence of TD. TD was diagnosed in 73 (8.4%) patients. The patients with TD had a higher mean daily neuroleptic dose than those without TD (386 ± 325 and 304 ± 245 mg/day chlorpromazine equivalent (CPZE). Although the overall daily dose of neuroleptic medications in this study was low, similar prevalence rates were reported from another patient population which had been prescribed much higher neuroleptic doses. Chiu et al. (25) reviewed 917 Chinese inpatients in Hong Kong who received a mean ± SD daily neuroleptic dose of 876 ± 853 mg CPZE. The overall prevalence grate for TD was 9.3%, and the rate for patients with schizophrenia was 8.4%.

Hungary. Gardos et al. (26) conducted a study of 122 patients with schizophrenia aged 15 to 66 (mean 37) years in Hungary and found a complete absence of severe TD. Although elderly women were underrepresented and there were several differences in prescribing practices between the Hungarian physicians and those in North America, the complete absence of severe TD in Hungary reported in this study was striking.

Studies in African Americans. Stramek et al. (27) conducted a study of chronic psychiatric inpatients and examined the effect of ethnicity on TD prevalence. The 491 patients examined included Caucasians (n = 280; 185 males, 95 females), African Americans (n = 112; 74 male, 38 female), and Hispanics (n = 99, 74 male, 25 female). The total study group ranged in age from 17 to 84 years (36.4 ± 10.6; mean ± SD). Using AIMS ratings, there were no significant differences in TD prevalence among the three groups. TD was present in 18.9% of the Caucasians, 17% of the African Americans, and 15.2% of the Hispanics for an overall prevalence rate of 17.7%. Mean daily neuroleptic doses (±SD) for all the ethnic groups were relatively high and were 1701± 1721, 2173 ± 1954, and 1785 ± 1708 mg CPZE, respectively, in patients without TD.

Results different from those of Sramek et al. (27) were reported by Holden (28) who conducted a prevalence study of TD among Black (Zulu) psychiatric patients in Natal, Republic of South Africa. Fifty men and 50 women were randomly selected from among 1000 long-term psychiatric inpatients and assessed by a rater who was blind to their diagnoses and medications. Patients were rated on the 4-point Kidger scale (29). The diagnoses patients received, which were obtained from the hospital psychiatrists, were schizophrenia (58%), epilepsy (24%), retardation (12%), and organic brain syndrome (6%); these were distributed evenly between the gender groups. Because of previous treatment policies, it was believed that most, if not all, of the patients had received electroconvulsive therapy at some time. Thirty-nine percent of the patients were found to have TD, although this prevalence might have been somewhat overestimated, according to the author, because spontaneous dyskinesia was probably present in some patients with chronic schizophrenia and organic brain syndrome. Twenty-six and 13 patients were found to have peripheral and orolingual TD, respectively. The grand mean global TD severity score was 3.8, indicating that most patients had dyskinesia in more than one body area (i.e lips, tongue, jaw, neck, trunk, arms, hands, or legs). These data were compared with those of a previous study by Holden et al. (30) which used similar methods with 287 Caucasian patients in Cape Town, Republic of South Arica. This group's primary diagnoses comprised schizophrenia (47%), organic brain syndrome with aberrant behavior (29%), manic-depressive illness (14%), and epilepsy with psychotic symptoms (10%). TD was seen in 49 (17%) of these patients. The grand mean global TD severity scores were nearly identical between the two studies (3.9 and 3.8 for the Zulus and Caucasians, respectively). The mean daily doses of neuroleptic medications in milligrams of CPZE, however, were markedly different (92 mg for Zulus and 381 mg for Caucasians). When this difference in the mean daily neuroleptic dose is taken into account, the higher TD prevalence rate found in Zulus (39% vs 17% in Caucasians) is strong evidence that the Zulus have a greater risk of developing TD than Caucasians.

Morgenstern and Glazer (31) reported a higher incidence (annual number of new cases) of TD among African Americans than among Caucasians. These authors examined patients from an outpatient mental health center who were receiving neuroleptic medications. A total of 398 patients without persistent TD entered the study and underwent follow-up visits every 6 months. At baseline 26% were nonwhite, and 97% of this group was African Americans. Patients were assessed for the incidence of persistent TD based on the AIMS (21) which was administered twice during each visit. Patient interviews continued for a total of 4 years, and only data from the patients who had been seen at least two times were analyzed. The relative risk of developing persistent TD for nonwhites was 1.81 (95% confidence interval 1.07–3.06) times that in the Caucasians ($p = 0.026$). This finding also remained when other demographic and clinical risk factors were controlled for. A more thorough analysis of the data regarding the relationship between ethnicity and TD was published later (32). The nonwhite patients were more likely to have received a diagnosis of schizophrenia than the white patients. The nonwhite subjects were also more likely to have received higher doses and higher potency neuroleptic medications during the study period than the white patients. When the use of depot versus oral medications was controlled for in addition to ethnicity and demographic characteristics, the difference in the mean neuroleptic dose between the whites and nonwhites was only 2.8 mg/day CPZE.

Furthermore, when ethnicity and any other study variable were controlled for, the rate ratio associated with ethnicity did not change more than 5%. This strongly suggested that the increased incidence of TD observed in the nonwhite patients was not caused by other variables but was related to potential differences in pharmacokinetics, pharmacodynamics, or neurophysiology between the whites and nonwhites in the study. Jeste et al. (22) examined 266 patients who were more than 45 years of age and were seen relatively early (average 21 days) in the course of their treatment with neuroleptic medications. The group had a mean ± SD age of 66 ± 12 years and included patients with schizophrenia (22%, mood disorders (21%), Alzheimer's disease (24%), and other neuropsychiatric disorders requiring treatment with neuroleptic medications (33%). Patients were usually treated with relatively low doses of haloperidol or thioridazine and received an average daily dose of 150 mg CPZE. After initial assessments, the patients were evaluated at 1 month, 3 months, and then every 3 months. Analysis derived from the Caucasian ($n = 215$) and African American ($n = 32$) patients had a cumulative annual incidence of TD of 26% in Caucasians and 48% in African Americans, although this difference did not reach statistical significance. The two ethnic groups did not differ significantly in daily neuroleptic dose or duration of treatment (33).

Summary. Asians probably have TD prevalence rates similar to or lower than those in Caucasians, while African Americans seem to have a somewhat higher risk of TD. Finally, to our knowledge there is no evidence of ethnicity as a risk factor for neuroleptic-induced parkinsonism.

Conclusion

Several early studies found lower prevalence rates of PD in African Americans than in Caucasians. However, more recent studies using superior surveying techniques did not find this difference, and it seems unlikely that there are any substantial differences in the prevalence or incidence of PD between ethnic groups. Of the limited number of TD studies examining race, it appears that African Americans may have a higher risk of developing TD than Caucasians. Deriving definitive interpretations from the studies is difficult, however, because there is usually a difference in the doses of neuroleptic medications used the different ethnic groups. Nonetheless, the relationship of ethnicity and TD deserves further investigation. Other movement disorders such as HD and neuroleptic-induced parkinsonism seem to have nearly equal rates among ethnic groups.

It would be premature to make definitive statements regarding the role of ethnicity in movement disorders based on the studies reviewed. Methodologic problems may result in both false positive and false negative results. Furthermore, there are issues related to an appropriate assessment of ethnicity, especially in a multiracial country such as the United States. If a significant relationship were demonstrated between ethnicity and a movement disorder, the association could be caused by racial-genetic factors and/or environmental factors such as socioeconomic status and access to adequate health care. Large, well-designed studies to clarify the role of ethnicity in movement disorders are warranted. A better understanding of the development and pathophysiology of movement disorders could lead to the development of more effective treatments and possibly even prevention strategies. For example, a higher risk of TD in African Americans may suggest the need for greater caution in the use of neuroleptic medications in that ethnic group.

References

1. Jeste DV, Caligiuri MP. Tardive dyskinesia. Schizophr Bull 1993; 19:303–315.
2. Kessler II. Epidemiological studies of Parkinson's disease II. A hospital-based survey. Am J Epidemiol 1972; 95: 308–318.
3. Kessler II. Epidemiological studies of Parkinson's disease III. A community-based survey. Am J Epidemiol 1972; 96:243–245.
4. Paddison RM, Griffith RP. Occurrence of Parkinson's dis-

ease in black patients at Charity Hospital in New Orleans. Neurology 1974; 24:688–690.

5. Lombard A, Gelfand M. Parkinson's disease in the African. S Afr Med J 1978; 24:5–8.
6. Jenkins AC. Epidemiology of parkinsonism in Victoria. Med J Aust 1966; 2:469–502.
7. Ashok PP, Radhakrishan K, Sridharan R, Mousa ME. Epidemiology of Parkinson's disease in Benghazi, North-East Libya. Clin Neurol Neurosurg 1986; 88:109–113.
8. Harada H, Nishikawa S, Takahashi K. Epidemiology of Parkinson's disease in Japanese city. Arch Neurol 1983; 40:151–154.
9. Li SC, Schoenberg BS, Wang CC, et al. A prevalence survey of Parkinson's disease and other movement disorders in the People's Republic of China. Arch Neurol 1985; 42:655–657.
10. Schoenberg BS, Anderson DW, Haerer AF. Prevalence of Parkinson's disease in the biracial population of Copiah County, Mississippi. Neurology 1985; 35:841–845.
11. Rajput AH. Epidemiology and etiology of Parkinson's disease. Washington DC: Office of Technology Assessment, 1988.
12. Schoenberg BS, Osuntokun BO, Adeuja AOG, et al. Comparison of the prevalence of Parkinson's disease in black populations in the rural United States and in rural Nigeria: Door-to-door community studies. Neurology 1988; 38:645–646.
13. Anderson DW, Schoenberg BS, Haerer AF. Prevalence surveys of neurologic disorders: Methodologic implications of the Copiah County study. J Clin Epidemiol 1988; 41:339–345.
14. Haddock RDW. Neurological disorders in Ghana. In: Spilane JD, editor. Tropical neurology. New York: Oxford University Press; 1973.
15. Wright HH, Still CN, Abramson RK. Huntington's disease in black kindreds in South Carolina. Arch Neurol 1981; 38:412–414.
16. Hayden MR, MacGregor JM, Saffer DS, Beighton PH. The high frequency of juvenile Huntington's chorea in South Africa. J Med Genet 1982; 19:94–97.
17. Folstein SE, Chase GA, Wahl WE, McDonnell AM, Folstein MF. Huntington disease in Maryland. Clinical aspects of racial variation. Am J Genet 1987; 41:168–179.
18. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Washington DC: American Psychiatric Association; 1994.
19. Yassa R, Jeste DV. Gender differences in tardive dyskinesia. A critical review of the literature. Schizophr Bull 1992; 18:701–715.
20. Hayashi T, Yamawaki S, Nishikawa T, Jeste DV. Use and side effects of neuroleptics in elderly Japanese patients. Am J Geriatr Psychiatry 1995; 3:308–316.
21. National Institute of Mental Health. Abnormal Involuntary Movement Scale (AIMS). Early Clinical Drug Evaluation Unit Intercom. 1975; 4:3–6.
22. Jeste DV, Caligiuri MP, Paulsen JS, et al. Risk of tardive dyskinesia in older patients. A prospective longitudinal study of 266 patients. Arch Gen Psychiatry 1995; 52: 756–765.
23. Ogita K, Yagi G, Itoh H. Comparative analysis of persistent dyskinesias of long-term usage with neuroleptics in France and in Japan. Folia Psychiatr Neurol Jpn 1975; 9:315–320.
24. Ko G, Zhang LD, Yan WW, et al. The Shanghai 800. Tardive dyskinesia prevalence in China. Am J Psychiatry 1989; 146:387–389.
25. Chiu H, Shum P, Lau J, et al. Prevalence of tardive dyskinesia, tardive dystonia, and respiratory dyskinesia among Chinese psychiatric patients in Hong Kong. Am J Psychiatry 1992; 49:1081–1085.
26. Gardos G, Samu I, Kallos M, Cole J. Absence of severe tardive dyskinesia in Hungarian schizophrenic out-patients. Psychopharmacology 1980; 71:29–34.
27. Sramek J, Roy S, Ahrens T, et al. Prevalence of tardive dyskinesia among three ethnic groups of chronic psychiatric patients. Hosp Community Psychiatry 1991; 42:590–592.
28. Holden TJ. Tardive dyskinesia in long-term hospitalized Zulu psychiatric patients. S Afr Med J 1987; 71:88–90.
29. Kidger T, Barnes TRE, Trauer T, Taylor PJ. Sub-syndromes of tardive dyskinesia. Psychol Med 1980; 10: 513.
30. Holden TJ, Sandler R, Myslobodsky M. Tardive dyskinesia prevalence and subtypes at Valkenberg Hospital, Cape Town. S Afr Med J 1984;66:132–134.
31. Morgenstern H, Glazer WM. Identifying risk factors for tardive dyskinesia among long-term outpatients maintained with neuroleptic medications: Results of the Yale Tardive Dyskinesia Study. Arch Gen Psychiatry 1993; 50:723–733.
32. Glazer WM, Morgenstern H, Doucette J. Race and tardive dyskinesia among outpatients at a CMHC. Hosp Community Psychiatry 1994; 45:38–42.
33. Lacro JP, Jeste DV. The role of ethnicity in the development of tardive dyskinesia. In: Nair NVP, Yassa R, Jeste DV, editors. Neuroleptic-induced movement disorders. London: Cambridge University Press; 1996, in press.

Ethnicity and Antidepressant Response

John J. Sramek, Pharm.D.[1] and Edmond H. Pi, M.D.[2]

Abstract

Modern psychopharmacology provides us with several effective antidepressants. We are increasingly recognizing the importance of interindividual differences in plasma concentration with these drugs. Such differences are caused by genetics, and the more extreme variations result from the population classified as slow hydroxylators. We now recognize that there are also substantial racial differences in the frequency distribution of slow metabolizers. The importance of studies in this area has been underscored by anecdotal reports as well as cross-cultural surveys of prescribing patterns which indicate that antidepressant dosage requirements may differ among racial groups. This review will focus on the relatively young field of cross-cultural psychopharmacology and the efforts being made to determine the importance of society, culture, environment, genetics, and biophysiology in the prescribing and metabolism of, and response to antidepressant drugs.

The primary goal of antidepressant therapy is to achieve remission of all depressive symptoms. Modern psychopharmacology provides us with several effective antidepressant agents, but treatment is often hampered by side effects, delayed onset of action, and lack of response in a significant fraction (20–30%) of patients with depression (1). Although efforts have been made to develop newer agents which are more effective or have more rapid onset of action than the tricyclic antidepressant drugs, none yet are ideal (2). On the surface, various antidepressant agents seem to have complex and diverse actions on neurotransmitters, although a common pathway of action has been hypothesized (3).

We are increasingly recognizing the importance of interindividual differences in plasma concentrations with these drugs, which can vary 30- to 40-fold with the tricyclic agents. Such differences in rates of metabolism are caused by genetics, with the more extreme variations resulting from the population classified as slow hydroxylators (4, 5). We now recognize that there are also substantial racial differences in the frequency distribution of slow metabolizers (6). The importance of studies in this area has been underscored by anecdotal reports as well as transcultural surveys of prescribing patterns which indicate that antidepressant dosage requirements may differ among racial groups (7–9).

The relatively young field of transcultural psychopharmacology has begun efforts to determine the relative importance of society, culture, environment, genetics, and biophysiology on the prescribing and metabolism of, and response to psychotherapeutic drugs. Cross-cultural studies are notoriously difficult to interpret because of the many inter- and intracultural variations in diagnostic criteria for mental illnesses, criteria for therapeutic responses, recommended dosages, attitudes toward health care, and incidence/severity of unpleasant side effects. Furthermore, some confusion arises from the tendency to group all members of a race together without regard to ethnic or cultural background. Clearly these features are interrelated; but whereas “race” is a classification which groups individuals who share distinct genetic characteristics, “ethnic groups” are subpopulations that are set apart based on any number of traits which may include religion, language, appearance, ancestry, culture, or nationality. Some ethnic groups have a homoge-

---

[1]From California Clinical Trials, Beverly Hills, California and [2]University of Southern California School of Medicine, Los Angeles, CA. Address reprint requests to John J. Sramek, Pharm.D., Director of Research, California Clinical Trials, 8500 Wilshire Blvd., 7th Floor, Beverly Hills, CA 90211.

neous gene pool; others do not. Culture, too, plays an important role that makes interpretation of transcultural research more complicated. Cultures differ in eating and food preparation habits, in exposures to pesticides, smog, and other toxins, and in caffeine, nicotine, and other drugs use patterns. All of these factors can affect the processing of a drug and must be considered when trying to determine whether or not there are group differences in drug disposition. With these caveats in mind, we will review the literature on racial/ethnic response to antidepressant medication and offer methodological suggestions for future studies.

Pharmacokinetics and Pharmacodynamics

Simply defined, pharmacokinetics (PK) can be described as what the body does to a drug (e.g., metabolize, eliminate, etc.), whereas pharmacodynamics (PD) is what the drug does to the body (e.g., beneficial effects and side effects) (10). Together, PK and PD comprise our understanding of the pharmacology of a class of compounds. The magnitude of a compound clinical effect is the product of PK and PD, but is also a product of individual and ethnic biological variances.

PK Differences in Ethnic Response to Antidepressant Medications

An early study by Allen et al. (11) reported on 6 East Asian (Indian/Pakistani) and 11 Caucasian male volunteers who received single doses of 25 mg or 50 mg clomipramine. Asians receiving 50 mg clomipramine attained significantly higher peak plasma and mean peak concentrations than Caucasian subjects. A later study expanded the Caucasian population to 25, with the same result (12). Like most of the studies reviewed here, this study is limited by the small sample size, which increases the chance that an existing difference will not be found.

In an oft-cited study, Rudorfer et al. (13) compared the clearance of 100 mg desipramine and its metabolite in 14 Chinese and 16 Caucasian volunteers. Most of the volunteers had intermediate rates of clearance regardless of ethnicity, but a few Chinese had slow clearances and a few Caucasians had fast clearances, leading investigators to propose the concept of a trimodal distribution in which Chinese have slow or intermediate rates of clearance and Caucasians have intermediate or rapid rates. The same investigators looked for a similar pattern using debrisoquine in 10 Chinese and 10 Caucasian volunteers, all but one of whom had participated in the previous study (14). Unlike desipramine, debrisoquine was cleared rapidly by every subject, including those who had shown slow clearance in the previous study. Because debrisoquine was thought to use the same metabolic pathway as desipramine and was therefore predicted to behave the same way, these results were difficult to explain. One strength of these two studies is that they used an Asian group made up entirely of Chinese, several of whom were genetically related. Use of more genetically homogeneous groups enhances the possibility of detecting differences caused by genetics. Use of a generic Caucasian population in the same studies, on the other hand, has a downside, because people who descended from unrelated Caucasian peoples are likely to be genetically dissimilar.

Kishimoto and Hollister (15) reported a small study of single doses of nortriptyline in 10 male Caucasians in the United States and 10 male Japanese in Japan. Caucasians received 100 mg nortriptyline in capsule form, whereas Japanese received 50 mg in powdered suspension. Although the Japanese received half the dose taken by the Caucasians, they reached somewhat higher peak plasma concentrations and had a significantly higher mean area under the curve, a finding which the investigators interpreted as a greater bioavailability of nortriptyline in Japanese. However, the differences in plasma concentration seem equally likely to be caused by formulation differences. Ideally, studies of this nature should use a single batch of study drug provided by the manufacturer. Furthermore, this study does not indicate whether or not a single laboratory was used in the PK analysis. Studies of populations in more than one location should strive to use a single laboratory to reduce variances arising from differences in analytical methods and technique. In a similarly designed study, Pi et al. (16) compared the PK of single doses of 50 mg desipramine in 20 Asians with 75 mg desipramine in 20 Caucasians. Asians were found to have a significantly earlier time to peak plasma concentration (4.0 hr vs 6.9 hr; $p < 0.05$). Despite having relatively large sample size for research of its kind, the authors noted that statistical confidence remained low, limiting the generalizability of this finding to larger populations. The same investigators undertook a second study (17) similar to that of Rudorfer (13), looking at the clearance of 1 mg/kg desipramine in 18 Asian and 19 Caucasian volunteers. Significant differences were found in time to peak plasma concentration (5.0 hr for Asians vs 3.0 hr for Caucasians; $p < 0.05$) and in desipramine and metabolite clearance, with Asians being slower than Caucasians in all cases. Adjusting for body weight made clearance differences nonsignificant. Nevertheless, the clearance patterns spurred investigators to suggest that a trimodal distribution does exist for desipramine clearance.

A recent study of the metabolism of clomipramine does not support this conclusion. Shimoda et al. (18) studied 108 Japanese psychiatric patients, more than 80% of whom were experiencing a major depressive episode. Clomipramine was titrated up to doses ranging from 30 mg/day to 250 mg/day (average 112 mg/day, or 2.18 mg/kg body weight), and patients were maintained at the same dose for 14 days before blood sampling was performed. At the conclusion of the analysis, no poor desmethylators, hydroxylators, or glucuronidators were found, providing no evidence that Asians might metabolize clomipramine more slowly than patients of other races. One limitation of most of the studies discussed so far [three exceptions being Allen et al. (11), Lewis et al. (12), and Kishimoto and Hollister (15)] is that both female and male volunteers were used. Gender-related factors such as menopausal status, hormone replacement therapy, menstrual cycle changes, etc. could affect the absorption, distribution, or metabolism of drugs such that within-group variability would increase, making it more difficult to detect between-group differences. Furthermore, in some of the studies, both Asian and Caucasian groups were composed of more than one racial group. For example, in Pi's studies, Asians could be Vietnamese, Chinese, Korean, or Japanese, and Caucasians came from various European groups.

Much less is known about the PK of antidepressants in people of African or Hispanic descent than of Asian descent. Ziegler and Briggs (19) treated 65 patients who were depressed with amitriptyline and nortriptyline and assessed the influence of age, race, sex, and smoking on plasma drug concentration. Black patients had significantly higher (50%) nortriptyline plasma levels than white patients, a finding which could explain why black patients may respond more quickly to tricyclic antidepressant medications. The study remains difficult to interpret, however, because diagnostic criteria are unclear, no gender distribution was reported, and, most importantly, a fixed- dose protocol was not used. In another area of research, although it has not been fully investigated, African American and African Black patients with bipolar disorder have been shown to have a higher red blood cell/plasma lithium ratio than Caucasian patients with bipolar disorder, which suggests that people of African descent might require lower doses of lithium to achieve antidepressant and/or antimanic efficacy (20). With regard to Hispanic response, no differences were found in the PK of a single dose of nortriptyline (75 mg) in one study of Mexican American and Caucasian volunteers (21). Investigators speculated that because depression itself may alter PK, the use of healthy volunteers may have decreased the value of findings.

PD Differences in Ethnic Response to Antidepressants

There have not been any studies specifically designed to asses PD factors that may affect response to antidepressants among different ethnic groups. Asian patients hospitalized with severe depression appear to respond clinically to lower combined concentrations of imipramine and desipramine (130 ng/ml) than are generally reported in North American and European studies (180–200 ng/ml) (8, 23), suggesting that tricyclic dosage requirements may be governed by variances in brain receptor responsivity among different ethnic groups. Furthermore, anecdotal reports suggest that Asians have a lower threshold for experiencing adverse drug reactions (7, 8). In the studies of Allen et al. (11) and Lewis et al. (12), East Asian subjects experienced significantly more drug-related side effects, and their reaction times were more affected by the medication than were the reaction times of Caucasians.

Several studies of lithium treatment for bipolar depression, furthermore, have tended to support the belief that Asians require lower dosages of therapeutic medication to achieve positive response. Yang (23) studied 101 Taiwanese patients with bipolar depression treated over 2 years with clinically determined doses of lithium, and found that plasma lithium levels for the majority of good responders ranged between 0.5 and 0.79 mEq/L. Similarly, Chang et al. (24, 25) reported optimal therapeutic lithium concentrations of 0.71 and 0.73 mEq/L for Chinese patients with bipolar depression living in Shanghai and Taipei, respectively, compared with a mean level of 0.98 mEq/L for matched Caucasian-American patients. These values are also lower than the 0.8 to 1.2 mEq/L generally thought to be standard in Europe and North America (26).

In a study of 159 black and 555 white patients with depression comparing chlorpromazine, imipramine, and placebo, black patients showed greater improvement at 1 week regardless of treatment, suggesting that blacks might be more sensitive to antidepressant effects (27). Furthermore, a chart review of 125 inpatients with depression (102 whites and 23 blacks) found that delirium was significantly more common in black patients and in those with higher tricyclic antidepressant levels (28). In another retrospective investigation conducted in 19 patients who had overdosed on amitriptyline, plasma concentrations were shown to be significantly higher in the six black patients (all female) than in the white patients (5 males and 13 females) (29). Although both these studies are limited by the small numbers of black patients or small size in general and by the lack of rigorous controls, they do suggest that African Americans are more susceptible to the central nervous system side effects of tricyclic antidepressant medications than are Caucasians. Whether or not this is caused by PD mechanisms is an issue which remains largely unexplored.

Some differences in response to antidepressants have been reported for Hispanic patients as well. Marcos and Cancro (30) studied 41 Hispanic (predominantly Puerto Rican) and 21 Caucasian women who were taking tricyclic antidepressant medications. The Hispanic women used half the dosage given to Caucasian women but had similar treatment outcomes. Furthermore, Hispanic patients complained of side effects and their medications were discontinued because of side effects significantly more often than Caucasians. Unfortunately, no PK data were collected in this study. Cultural factors relating to the significance of somatic complaints as depressive equivalents in the Hispanic population were thought to play an important role in the response to treatment.

Recommendations for Research

If differences in antidepressant drug disposition do exist among different ethnic/racial groups, an understanding of these differences could be a tremendous aid in facilitating psychiatric treatment of diverse populations, particularly in countries with heterogeneous populations such as the United States. To date, most transcultural psychopharmacology studies have been limited by their small and often inadequately characterized samples. Future studies should use larger populations to permit detection of between-group differences with statistical confidence. Populations need to be extremely well characterized to maximize the ability to control for other variables, including gender, age, lean and actual body weight, birthplace, residence, concomitant medical illnesses and/or medications, diet, alcohol/nicotine/caffeine intake patterns, menopausal status, menstrual cycle effects, and, of course, diagnosis and severity of depression (31). If a sample encompasses patients of several races residing in different countries, it may become easier to separate the influences of culture, race, and ethnicity.

At the same time, such a study would need to be monitored carefully to minimize the problems inherent in large, multicenter studies. To the greatest extent possible diagnostic criteria and rating and diagnostic techniques should be standardized. Furthermore, medication to be tested should be from a single source and given in a single dosage form to minimize variances that arise from product formulations. Future research into the PK of antidepressant medications among ethnic groups should also be expanded to include study of the serotonin-selective reuptake inhibitors. This class of drugs is widely prescribed for the treatment of depression, but the possibility of ethnic variations in response has not been systematically considered. At this time, the anecdotal evidence for PK and PD differences in various ethnic groups has gained limited support, at best, and more research should be performed. Systematic exploration of biological markers for depression in different populations is highly recommended. Using the dexamethasone suppression test, Escobar (32) found significant differences in cortisol suppression across groups of outpatients with depression at a Veterans Administration hospital: Fifty-eight percent of Anglo patients with a major depressive disorder were nonsuppressors compared with only 25% of blacks and 0% of Hispanics. A later study of 553 nonmedicated Mexican outpatients (33) concluded that there is little evidence to suggest differences among groups on the dexamethasone suppression test, although the proportion of dexamethasone suppression test nonsuppression among the 30 patients in the sample with major depressive disorders of 36% is still significantly different from the 58% figure for Anglo patients cited by Escobar. Clearly, elucidation of causes of differing values among different groups, if such differences are shown to exist, could be valuable.

Future study of other biological markers in patients with depression should examine levels of serotonin and norepinephrine metabolites in cerebrospinal fluid, plasma adrenocorticotropic hormone, and growth hormone, and compare platelet imipramine binding across groups. A few limited studies (e.g., 34–36) have begun to examine the possibility that enzyme activities differ among ethnic groups, leading to differences in PK and PD. This issue should be investigated further through genetic and environmental analyses to determine the degree of influence of each factor. In addition, more direct imaging methods such as positron emission tomography may enable investigators to look at specifically targeted areas of the brain. Systematic investigation of the potential effects of ethnicity on the biological correlates of depression may help to improve our understanding and treatment of depression and other mental disorders. Although we devote significant efforts to the research of biological markers, we must also establish that clinically significant variations in response to antidepressants do indeed exist between different ethnic groups with respect to efficacy or side effects.

Acknowledgments

The authors thank Giedra Miller for her assistance in writing this manuscript.

References

1. Dawkins K, Manji HK, Potter WZ. Pharmacodynamics of antidepressants. In: Cutler NR, Sramek JJ, Narang PK, editors. Pharmacodynamics and drug development: perspectives in clinical pharmacology. Chichester, England: John Wiley & Sons Ltd, 1994:157–180.
2. Katz MM, Koslow SH, Maas JW, et al. The timing, specificity, and clinical prediction of tricyclic drug effects in depression. Psychiatry Med 1987; 17:297–309.
3. Potter WZ, Rudorfer MV, Linnoila M. New clinical studies support a role of norepinephrine antidepressant action. In: Barchas JD, Bunney Jr. WE, editors. Perspectives in psychopharmacology. New York: Alan R Liss; 1988: 495–513.
4. Bertilsson L, Mellstršm B, Nordin C, et al. In: Gram LF, Usdin E, Dahl SG, et al. editors. Clinical pharmacology in psychiatry: bridging the experimental-therapeutic gap. London: Macmillan Press, 1983:217–226.
5. Brisen K, Otton SV, Gram LF. Sparteine oxidation polymorphism in Denmark. Acta Pharmacol Toxicol 1985; 57:357–360.
6. Meyer OA. Molecular genetics and the future of pharmacokinetics. In: Kalow W, editor. Pharmacogenetics of drug metabolism. New York: Pergamon Press; 1992.
7. Yamamoto J, Fung D, Lo S, Reece S. Psychopharmacology for Asian Americans and Pacific Islanders. Psychopharmacol Bull 1979; 15:29–31.
8. Yamashita I, Asano Y. Tricyclic antidepressants: therapeutic plasma level. Psychopharmacol Bull 1979; 15: 40–44.
9. Pi EH, Jain A, Simpson GM. Review and survey of different prescribing practice in Asia. In: Shagass C, Josiassen RC, Bridger W, Weiss KJ, editors. Biological psychiatry. New York: Elsevier Science Inc 1986:1536–1538.
10. Benet LZ. Foreward. In: Cutler NR, Sramek JJ, Narang PK, editors. Pharmacodynamics and drug development: perspectives in clinical pharmacology. Chichester, England: John Wiley & Sons Ltd; 1994: xiii–xv.
11. Allen JJ, Rack PH, Vaddadi KS. Differences in the effects of clomipramine on English and Asian volunteers: preliminary report on a pilot study. Postgrad Med J 1977; 53(S4):79–86.
12. Lewis P, Rack PH, Vaddadi KS, Allen JJ. Ethnic differences in drug response. Postgrad Med J 1980; 56(S1): 46–49.
13. Rudorfer MV, Lane EA, Chang WH, et al. Desipramine pharmacokinetics in Chinese and Caucasian volunteers. Br J Clin Pharmacol 1984; 17:433–440.
14. Rudorfer MV, Lane EA, Potter WZ. Interethnic dissociation between debrisoquine and desipramine hydroxylation. J Clin Psychopharmacol 1985; 5:89–92.
15. Kishimoto A, Hollister LE. Nortriptyline kinetics in Japanese and Americans. J Clin Psychopharmacol 1984; 4:171–172.
16. Pi EH, Simpson GM, Cooper TB. Pharmacokinetics of desipramine in Caucasian and Asian volunteers. Am J Psychiatry 1986; 143:1174–1176.
17. Pi EH, Tran-Johnson TK, Walker NR, et al. Pharmacokinetics of desipramine in Asian and Caucasian volunteers. Psychopharmacol Bull 1989; 25:483–487.
18. Shimoda K, Noguchi T, Ozeki Y, et al. Metabolism of clomipramine in a Japanese psychiatric population: hydroxylation, desmethylation, and glucuronidation. Neuropsychopharmacology 1995; 12:323–333.
19. Ziegler VE, Briggs JT. Tricyclic plasma levels—effect of age, race, sex, and smoking. JAMA 1977; 438:2167–2169.
20. Strickland T, Lin KM, Fu P, et al. Comparison of lithium ratio between African-American and Caucasian bipolar patients. Biol Psychiatry 1995; 37:325–330.
21. Gaviria M, Gil AA, Javaid JI. Nortriptyline kinetics in Hispanic and Anglo subjects. J Clin Psychopharmacol 1986; 6:227–231.
22. Glassman AH, Perel JM, Shostak M, et al. Clinical implications of imipramine plasma levels for depressive illness. Arch Gen Psychiatry 1977; 34:197–204.
23. Yang YY. Prophylactic efficacy of lithium and its effective plasma levels in Chinese bipolar patients. Acta Psychiatr Scand 1985; 71:171–175.
24. Chang SS, Pandey GN, Zhang MY, et al. Racial differences in plasma and RBC lithium levels. Presented at the Annual Meeting of the American Psychiatric Association, May 5–11, Los Angeles, CA. Continuing Medical Education Syllabus and Scientific Proceedings 1984:239–240.
25. Chang SS, Pandey GN, Yang YY, et al. Lithium pharmacokinteics: interracial comparison. Presented at the 138th Annual Meeting of the American Psychiatric Association; 1985 May 19–24; Dallas, Texas.
26. Jefferson JW, Greist JH, Ackerman DL, Carrol JA, editors. Lithium encyclopedia for clinical practice. 2nd ed. Washington DC: American Psychiatric Press; 1987.
27. Raskin A, Crook TH. Antidepressants in black and white inpatients. Arch Gen Psychiatry 1975; 32:643–649.
28. Livingston RL, Zucker DK, Isenberg K, Wetzel RD. Tricyclic antidepressants and delirium. J Clin Psychiatry 1983; 44:173–176.
29. Rudorfer MV, Robins E. Amitriptyline overdose: clinical effects of tricyclic antidepressant plasma levels. J Clin Psychiatry 1982; 42:457–460.
30. Marcos LR, Cancro R. Pharmacotherapy of Hispanic depressed patients: clinical observations. Am J Psychotherapy 1982; 36:505–513.
31. Pi EH, Wang AL, Gray GE. Asian/non-Asian transcultural tricyclic antidepressant psychopharmacology: a review. Prog Neuropsychopharmacol Biol Psychiatry 1993; 17:691–702.
32. Escobar JI. Are results on the dexamethasone suppression test affected by ethnic background? (letter). Am J Psychiatry 1985; 142:268.
33. de la Fuente JR, Sepulveda Amor J. Does ethnicity affect

DST results? (letter). Am J Psychiatry 1986; 143:275–276.

34. Bertilsson L, Lou YQ, Du YL, et al. Pronounced differences between native Chinease and Swedish populations in the polymorphic hydroxylations of debrisoquine and *S*-mephenytoin. Clin Pharmacol Ther 1992;51:388–397.
35. Nakamura K, Goto R, Ray WA, et al. Interethnic differences in generic polymorphism of debrisoquine and mephenytoin hydroxylation between Japanese and Caucasian populations. Clin Pharmacol Ther 1985; 38:402–408.
36. Horai Y, Nakano M, Ishizaki T, et al. Metoprolol and mephenytoin oxidation polymorphisms in Far Eastern Oriental subjects: Japanese vs. mainland Chinese. Clin Pharmacol Ther 1989; 46:198–207.

Gender Differences in Psychotropic Medications

FREDA LEWIS-HALL, M.D.

Abstract

During the past decade pharmacological research has greatly enhanced the understanding of several variables affecting the prescription of psychotropic medication. One variable of increasingly recognized importance is the gender of the patient. Significant gender differences have been described for psychiatric disease prevalence, symptom presentation, treatment-seeking behavior, and receipt of psychotropic medication. Despite these findings little has been investigated in the potential gender differences in efficacy, dosage, and side effects influenced by pharmacokinetic or pharmacodynamic properties of psychotropic medications. This leaves an important gap in information of potential clinical importance. This relatively young field of study has a number of variables to consider including the impact of psychosocial issues, environmental components, and genetic and hormonal influences. We will review literature on the impact of gender on psychotropic medications and make recommendations for further research.

DESPITE SIGNIFICANT GENDER differences in psychiatric disease prevalence, symptom presentation, treatment-seeking behavior, and receipt of psychotropic medication, little research has been conducted regarding the potential gender differences in efficacy, dosage, and side effects influenced by pharmacokinetic or pharmacodynamic properties of psychotropic medications. This paper will briefly review the literature on gender differences in psychotropic medications.

Antipsychotic Agents

The literature analyzing potential gender differences is more substantial for antipsychotic agents than for other psychotropic groups. Four major areas have been investigated: overall efficacy, dose requirements, plasma levels, and side effects. Chouinard and Annable (1) evaluated efficacy with antipsychotic medications in a small population and noted a greater improvement for women treated with pimozide and chlorpromazine than for men. The authors hypothesized that this differential response was caused by the presumed antidopaminergic effect of estrogen (2, 3). Other authors have suggested that differences in dose requirements of women for antipsychotic medications may be mediated by the presumed antidopaminergic effect of estrogen. Chouinard et al. (4) examined the dosage requirements of fluspirilene, a long-acting injectable neuroleptic agent. They noted that women required significantly lower doses (nearly half) than men despite similarities in weight and age. Using a survey-type design, Seeman (5) also examined dosage requirements for neuroleptic agents (reported in chlorpromazine equivalents) for women. Women in their 20s and 30s were reported as requiring lower doses of antipsychotropic medications. Similar results were reported in a retrospective review of uncontrolled treatments by DiMello and MacNeil (6). Younger men and older women were reported as having received higher doses.

Several significant questions arose in these studies. The assumption that the prescribed dose represented the therapeutic dose in these studies is significant, but questionable, because no con-

---

From Lilly Research Laboratories, Indianapolis, Indiana. Address reprint requests to Freda Lewis-Hall, M.D., Director of Womens' Studies, Lilly Research Laboratories, Lilly Corporate Center, Indianapolis, IN 46225.

sistent outcome variables were described. Further, the diagnoses varied within studies and between studies. Effective doses were recognized as varying across diagnoses, further obscuring the accuracy of the therapeutic dose. Unfortunately, plasma levels were not available to correlate differences with therapeutic responses, but some studies have examined plasma levels of antipsychotics agents. Simpson et al. (7) examined levels of fluphenazine in women and found higher plasma levels in women when no significant differences in milligrams per kilogram dose were noted. Ereshetsky et al. (8) also noted higher levels of thiothixene in women. In addition to higher plasma levels, women were noted to have significantly lower clearance than men. Differences in antipsychotic-induced side effects have also supported the possibility of gender differences in these medications. Three side effects have been noted to be variably expressed by gender: tardive dyskinesia, Parkinson's disorder, and hematologic events. The incidence of tardive dyskinesia (TD) has been reported as greater in women than in men in several studies (9–11), and women have also been reported to suffer more severe TD. Smith et al. (11) reported that the severity of TD increased significantly in women older than 67 years, and a gender-age relationship was further described by Chouninard et al. (12) who noted that young men reported a higher prevalence of severe dyskinesias than women. One potential explanation was that postmenopausal status increases exposure to the development of tardive dyskinesia (13). This would be consistent with hypotheses that estrogen acts at the dopamine receptor, providing protection to premenopausal women. Yassar and Jeste (13) combined data from independent studies on prevalence, age, and sex differences in TD. Results indicate that TD is more prevalent in women and in old age. The authors raised several points which suggested that these findings are complicated by differences in diagnosis and length of exposure to neuroleptic medications. The impact on gender in the development of Parkinsonism has been evaluated infrequently. Jeste (14), in a close examination of rates and severity of movement disorders, recently reported that with lower doses of neuroleptic medications the risks of Parkinsonism in women remain similar to or lower than those in men. Hematologic events associated with treatment with newer neuroleptic therapies have been important in the clinical management of patients. Careful monitoring has provided some significant gender relationship differences. Banov et al. (15) reported that the incidence of eosinophilia in women treated with clozapine was significantly higher than in men. He also noted that eosinophilia occurred early in therapy, resolved spontaneously, and was not associated with known clinical complications. Although sparse, the available literature in this area supports gender-related findings which suggest a need for a further, more aggressive research agenda. The critical interaction between age and gender must be examined more closely. Additionally, menstrual status, menstrual cycle effects, and exogenous hormone administration need careful definition in studies.

Antidepressants

Because of the higher overall prevalence of depressive disorders in women, much more research is needed (16). Most studies in this area focus on efficacy, dosage, and plasma levels as variables affecting gender differences. Overall efficacy of antidepressant medications in women has been evaluated in several studies. Davidson and Pelton (17) reported that women with depression who have panic attacks have greater response to monoamine oxidase inhibitors than to tricyclic antidepressant agents. Men who had depression and panic attacks, however, had a more favorable response to tricyclic antidepressant agents. Raskin (18) noted that women (older than 40) and men had a better response to imipramine than did younger women. Young men also responded better to chlorpromazine. Younger women in the study responded better to phenelzine than younger men. This study highlighted the importance of the gender-age relationship also seen in antipsychotic medications. Of note in this retrospective analysis: the subjects did not necessarily have a diagnosis of depression, evidence for significant interdiagnosis differences in dose requirements exists, and failure to standardize diagnoses and further specify diagnostic criteria represent issues for future study.

Dose requirements and plasma levels have been examined in several studies, and the findings have been inconsistent. Burrows et al. (19) found no significant gender differences with nortriptyline. Zeigler and Biggs (20) noted similar findings with nortriptyline, and they also examined plasma levels in a small sample of subjects given amitriptyline and noted no gender differences. Preskorn and Mac (21) examined 110 patients in a naturalistic study of inpatients and noted that women had higher plasma levels per milligram of drug than men. Moody et al. (22) noted nonsignificantly higher plasma levels of imipramine in women patients. In the area of newer antidepressant medications, Greenblatt et al. (23) investigated gender differences in trazadone and reported that the volume of distribution was greater in women and clearance was reduced in older men. Several factors contributed to the variability of the findings: several studies had retrospective and "naturalistic" study designs which failed to account for diagnosis or therapeutic dose by a specific metric; small sample size and failure to control for variables such as coadministration of medication, smoking, hormone replacement therapy, or hormone status may have had an impact on outcomes; and several studies supported a relationship between plasma level and coadministration of oral contraceptives. Abernathy et al. (24) noted that oral contraceptives inhibited the hepatic metabolism of intravenous imipramine. This demonstrated the potential impact of oral contraceptives and other hormonal products on the metabolism of antidepressant medications (25).

Lithium is one of the few psychotropic medications that has been the subject of reports related to menstrual cycle effect. In two case reports (26, 27) higher lithium levels in women in postmenstrual versus premenstrual phases of the menstrual cycle were noted. Volunteers without bipolar disorder studied by Chamberlain et al. (27) failed to show a difference between mean lithium levels when lithium was given during the follicular and luteal phases. Of note, several individuals evidenced phasic differences. Individual reports by women with bipolar disorder of phasic changes in symptoms may suggest those "at risk" for cycle-dependent differences, and this warrants further study.

Antianxiety Agents

Much of the research on antipsychotic and antidepressant agents has been in efficacy, dosage, and side effects, whereas little has been done regarding pharmacokinetic and pharmacodynamic differences. Conversely, in antianxiety agents little research has been conducted in the area of therapeutic affects and closer attention has been paid to kinetics. This series of studies has given greater consideration to metabolic pathway as an additional factor to consider in evaluating gender differences. Greenblatt et al. (28) found that diazepam, a benzodiazepine metabolized through oxidation, had a higher clearance in younger women than in men. This study also examined the gender-age relationship and noted that the statistical significance of this finding disappeared in women in the 62- to 84-year age group. This study used a single intravenous dose, but did not assess metabolites. In a similar study, MacLeod et al. (29) noted no significant impact of age, finding that women metabolized diazepam slower than men regardless of age. Triazolam, also oxidatively metabolized, was found to evidence no gender differences in metabolism (30). Nitrazepam, metabolized by reduction, was not found to have any gender-related differences in clearance (31). Divoll et al. (32) examined the elimination half-life of temazepam, metabolized by glucuronidation, and found it be longer in women than men after single dose administration. Oxazepam, also metabolized by glucuronidation, was noted by Greenblatt et al. (33) to have a longer elimination half-life in women than in men. Half-life increased with age in women, but did not seem to be associated with age in men.

Several studies in benzodiazapines have examined the impact of the menstrual cycle and exogenous hormone administration metabolism. Giles et al. (34) reported that the clearance of diazepam was lower in oral contraceptive users than in control subjects. Ellinwood et al. (35) found that the greatest psychomotor and cognitive impairment occurred on day 28 of the menstrual cycle in women taking oral contraceptives. They postulated the oral contraceptives decrease the rate absorption of diazepam and during the week off hormones the plasma levels quickly rose to become more intoxicating. Jochemsen et al. (31) noted that women taking contraceptives and nitrazepam had higher steady-state levels than either men or women not taking oral contraceptives. Future studies in this area will require control gender-related effects on enzymatic function as some differences may depend on metabolic pathways. There also seems to be a great degree of interindividual variability, which implies a need to reduce reliance on reporting group findings in which individual differences may provide important information on how differences are mediated.

Conclusion

This paper has reviewed several studies which suggest that gender-related differences have an impact on clinically relevant parameters in psychotropic medication action. Recently, Dawkins and Potter (36) reviewed the substantial existing animal literature which supports the significant gender differences in animal models in both metabolism and response. Together, these two sources of information provide both provocation and caution to this field of research. The studies to date have generally been in small, poorly characterized populations. Future studies will require larger populations and those characterized by age, psychosocial factors, body weight, hormonal status, presence of exogenous hormones, concomitant illnesses and medications, and nutritional status (including alcohol and tobacco use). Further, stabilization of parameters used in diagnosis, therapeutic affect, and adverse events are required. The diversity and complexity of psychotropic medications have made it important to recognize that information on individual psychotropic agents can not be generalized. Future studies should focus on amassing sufficient data across variables on individual agents.

References

1. Chouinard G, Annable I. Pimozide in the treatment of newly schizophrenic patients. Psychopharmacology (Berl) 1982; 76:13–19.
2. Fields IZ, Gordon JH. Estrogen inhibits the dopaminergic supersensitivity induced by neuroleptics. Life Sci 1982; 30:229–234.
3. Villeneuve A, Langlier P, Bedard P. Estrogens, dopamine and dyskinesias. Can Psychiatr Assoc J 1978; 23:68–70.
4. Chouinard G, Annable I, Steinberg S. A controlled clinical trial of fluspirilene, a long-acting injectable neuroleptic in schizophrenic patients with acute exacerbation. J Clin Psychopharmacol 1986; 6:21–26.
5. Seeman MV. Interaction of sex, age and neuroleptic dose. Compr Psychiatry 1983; 24:125–128.
6. DiMello DA, MacNeil JA. Sex differences in bipolar affective disorder; neuroleptic dosage variance. Compr Psychiatry 1990; 31:80–83.
7. Simpson GM, Yakalam KG, Levinson DP, et al. Single dose pharmacokinetics of fluphenazine after fluphenazine decanoate administration. J Clin Psychopharmacol 1990; 10:417–421.
8. Ereshetsky L, Saklad SR, Watanabe MD, et al. Thiothixene-pharmacokinetic interactions: a study of hepatic enzyme inducers, clearance inhibitors, and demographic variables. J Clin Psychopharmacol 1991; 11: 269–301.
9. Smith JM, Dunn DD. Sex differences in the prevalence of severe tardive dyskinesia. Am J Psychiatry 1979; 136: 1080–1082.
10. Chouinard G, Annable I, Ross-Chouinard A, Nestoros JN. Factors related to tardive dyskinesia. Am J Psychiatry 1979; 136:79–83.
11. Smith JM, Kucharski IT, Oswald WT. Tardive dyskinesia: effect of age, sex, and criterion level of symptomatology on prevalence estimates. Psychopharmacol Bull 1979; 5:69–71.
12. Chouinard G, Jones BD, Annable I, Ross-Chouinard A. Sex differences and tardive dyskinesia (letter). Am J Psychiatry 1980; 137:507.
13. Yassar, Jeste DV. Gender differences in tardive dyskinesia: a critical review of the literature. Psychol Bull 1992; 1814:701–715.
14. Jeste DV. Gender and ethnicity differences in pharmacology of neuroleptics. American Psychiatric Association Annual Meeting 1995.
15. Banov MD, Tohen M, Friedberg J. High risk of eosinophilia in women treated with clozapine. J Clin Psychiatry 1993; 54:466–469.
16. Weissman MM, Klerman GL. Sex differences and the epidemiology of depression. Arch Gen Psychiatry 1977; 34:98–111.
17. Davidson J, Pelton S. Forms of atpyical depression and their response to antidepressant drugs. Psychiatr Res 1986; 17:87–95.
18. Raskin A. Age-sex differences in response to antidepressant drugs. J Nerv Ment Dis 1974; 159:120–130.
19. Burrows G, Scoggins BA, Turecek IR, Davies B. Plasma nortriptyline and clinical response. Clin Pharmacol Ther 1974; 16:639–644.
20. Zeigler VI, Biggs JT. Tricyclic plasma levels: effect of age, race, sex and smoking. J Am Med Assoc 1977; 238: 2167–2169.
21. Preskorn SH, Mac DS. Plasma level of amitriptyline: effect of age and sex. J Clin Psychiatry 1985; 46:276–277.
22. Moody JP, Tait AC, Todrick A. Plasma levels of imipramine and desmethylimipramine during therapy. Br J Psychiatry 1967; 113:183–193.
23. Greenblatt DJ, Friedman H, Bernstein ES, et al. Trazodone kinetics: effect of age, gender and obesity. Clin Pharmacol Ther 1987; 42:193–200.
24. Abernathy DR, Greenblatt DJ, Shader RI. Imipramine disposition in users of oral contraceptive steroids. Clin Pharmacol Ther 1984; 35:792–797.
25. Conrad CD, Hamilton JA. Recurrent premenstrual decline in serum lithium concentration. J Am Acad Child Psychiatry 1986; 26:852–853.
26. Kukopulos A, Minnai G, Muller-Oelinghausen B. The influence of mania and depression on the pharmacokinetics of lithium: a longitudinal single-case study. J Affect Disord 1985; 8:159–166.
27. Chamberlain S, Hahn PM, Cassan P, Reid RL. Effect of menstrual cycle phase and oral contraceptive use on serum lithium levels after a loading dose of lithium in normal women. Am J Psychiatry 1990; 147:907–909.
28. Greenblatt DJ, Allen MD, Harmatz JS, Shader RI. Diazepam disposition determinants. Clin Pharmacol Ther 1980; 27:301–312.
29. Macleod SM, Giles HG, Bengerr B, et al. Age and gender related differences in diazepam pharmacokinetics. J Clin Pharmacol 1979; 19:15–19.
30. Smith RB, Divoll M, Gillespie WR, Greenblatt DJ. Effect of subject age and gender on the pharmacokinetics of oral triazolam and remazepam. J Clin Psychopharmacol 1983; 3:172–176.
31. Jochemsen R, Van der Graaff M, Boeijinga JK, Breimel D. Influence of sex, menstrual cycle and oral contraception on the disposition of nitrazepam. Br J Clin Pharmacol 1982; 13:319–324.
32. Divoll M, Greenblatt DJ, Harmatz JS, Shader RI. Effect of age and gender on disposition of temazepam. J Pharm Sci 1981; 70:1104–1107.
33. Greenblatt DJ, Divoll M, Harmatz JS, Shader RI. Oxazepam kinetics: effects of age and sex. J Pharmacol Exp Ther 1980; 215:86–91.
34. Giles HG, Sellers EM, Naranjo CA, Greenblatt DJ. Disposition of intravenous diazepam in young men and women. Eur J Clin Pharmacol 1981; 20:207–213.
35. Ellinwood EH, Easier MP, Linnoila M, et al. Effects of oral contraceptives and diazepam-induced psychomotor impairment. Clin Pharmacol Ther 1983; 35:360–366.
36. Dawkins K, Potter WZ. Gender differences pharmacokinetics and pharmacodynamics of psychotropics: focus on women. Psychopharmacol Bull 1991; 27:417–426.

Ethnicity and Psychopharmacology in Latin America

Luis F. Ramirez, M.D.

Abstract

It has been reported anecdotically that patients from Latin America need less medication than Anglo patients for the treatment of psychiatric disorders. These clinical observations have been questioned because of the character of the reports. However, ethnic differences in pharmacokinetics and pharmacodynamics are known to cause clinically observed differences in the effects of certain drugs. After investigation, such effects have been found to be pronounced in Asians, but few studies have been conducted with Hispanics. Fortunately, several studies are underway in Latin America, as reviewed in this brief report.

It has been reported anecdotically that patients from Latin American countries need less medication than Anglo patients for the treatment of disorders such as major depression, panic, or schizophrenia. These clinical observations have been questioned, but clinicians in Latin American countries do use smaller doses of psychotropic medications than the amounts used by their colleagues in North America. Ethnic differences in pharmacokinetics and pharmacodynamics are known to cause clinically observed differences in the effects of certain drugs, like alcohol or beta-blockers, in different populations. These effects have been examined and found to be more pronounced in Asian than in Latino populations, but few studies have been carried out in Latin America. Fortunately several efforts are underway to study the importance of ethnicity in psychopharmacology.

Pharmacokinetics and Pharmacodynamics

Concomitant with these efforts, significant advances have been made in the study of pharmacokinetics and pharmacodynamics in different populations. For example, two of the most frequently encountered P-450 isoenzymes, the CYP2D6, responsible for the metabolism of tricyclic antidepressant, selective serotonin reuptake inhibitor, and antipsychotic agents, such as clozapine and risperidone, and the YCYP2C19 responsible for the metabolism of some benzodiazepines are polymorphic in nature and have cross-ethnic variations. In terms of practical issues this polymorphism is manifested in some patients (who possess normally functioning enzymes) are poor metabolizers (PMs) or extensive metabolizers (EMs). The PM rate for CYP2D6 ranges from 0.8% for Asians, 1.9% for African Americans, 4.5% for Hispanics to 7% for Caucasians. The rates for CYP2C19 are 20% for Asians, 18.5% for African Americans, 4.8% for Hispanics, and 3.3% for Caucasians (1). Another ethnic factor of importance, body size and percentage of fat, influences the volume of distribution of lipophilic psychotropic agents. These differences in body size and fat distribution may explain some of the differences in doses between Latino and Anglo pa-

---

From Case Western Reserve University and Cleveland Veterans Administration Medical Center. Address reprint requests to Luis F. Ramirez, M.D., Associate Professor and Vice-Chairman, Department of Psychiatry, Case Western Reserve University and Chief of Psychiatry Services, Cleveland Veterans Administration Medical Center, 10000 Brecksville Road, Brecksville, OH 44141.

tients because of the high lipophilic profile of many psychotropic agents.

Typical and Atypical Antipsychotics

Several studies have investigated plasma concentrations of medications in Asian populations. Healthy Asian volunteers and Asian patients with schizophrenia had approximately 50% higher plasma haloperidol concentrations than their Caucasian counterparts. These differences cannot be explained at the present time (2). A recent review of literature conducted by the author showed no published studies looking at the blood levels of antipsychotic medications in Latin Americans who are healthy and/or who have schizophrenia. Sandoz Pharmaceuticals in Argentina is actively studying oral doses, blood levels, clinical response, and quality of life in patients treated with clozapine. The hypothesis is that the Argentinean patients will require less oral medication to achieve the therapeutic blood levels of clozapine described in the literature by Potkin et al. (5). On the other hand, Gaviria et al. (3) have reported that the pharmacokinetics of nortriptyline in Hispanics are similar to the kinetics in Anglos, which casts some doubt on the hypothesis described above.

A study presented by Susan Viale-Hansen, from Argentina, during the Pan American Congress of Psychiatry in Mexico City in 1994 showed that in 60 patients treated successfully with clozapine the dose range was 25 to 600 mg per day during the acute phases, with a maintenance average dose of 300 mg per day. Similar results from a larger sample were presented by Veronica Larach from Chile. These dosages are smaller than the amounts used in North America, as reported by Meltzer and his group (6). Finally, in a consensus meeting organized by Janssen Pharmaceuticals during the first reunion of the Latin American College of Neuropsychopharmacology in Brazil, psychiatrists from Colombia, Argentina, and Mexico reported that the dosages of risperidone used to treat their patients with schizophrenia was similar to the amounts recommended by Janssen in the United States (6 mg daily), but with a tendency toward the use of smaller dosages in a large number of patients.

Comments

There is a general belief, from Mexico to Argentina, that Latin American patients need less medication than their North American counterparts, a generalization not confirmed by published reports. It is important for us to realize that Latin Americans are not a homogeneous population but a melting pot of many different cultures, and many "nonbiological" factors influence treatment outcome. Factors such as availability of care, biases on the part of health workers, special beliefs of patients, expectations, placebo effects, economics, and compliance are as important as the efficacy of the medication. For example Escobar and Tuason (7) reported that non-Caucasians may respond better to placebo than Caucasians. The World Health Organization has also reported that the outcome of patients with schizophrenia is better in underdeveloped than in developed countries (8). These two examples and others make the information available about dosage and clinical response in Latin American patients difficult to interpret. Consequently, more epidemiological and cross-cultural studies must be conducted to determine whether these clinical observations, that less medication is needed to achieve remission in patients with Latin American ancestry, are accurate.

References

1. Lin K-M, Anderson D, Poland RE. Ethnicity and psychopharmacology. Psychiatr Clin North Am 1995; 18:3.
2. Potkin SG, Shen Y, Pardes H, et al. Haloperidol concentrations elevated in Chinese patients. Psychiatr Res 1984; 12:167–172.
3. Gaviria MA, Gill AA, Javaid JI. Nortriptyline kinetics in Hispanic and Anglo subjects. J Clin Psychopharmacol 1986; 6:227–231.
4. Leon CA. Therapeutic effects of clozapine. A 4-year follow-up of a controlled clinical trial. Acta Psychiatr Scand 1979; 59:471–480.
5. Potkin SG, Bera R, Gulasekaram B, et al. Plasma clozapine concentrations predict clinical response in treatment-resistant schizophrenia. J Clin Psychiatry 1994; 55:133–137.
6. Meltzer HY, Alphs L, Bastani B, et al. One year outcome of study of clozapine in treatment resistant schizophrenia. Presented at 8th World Congress of Psychiatry; Athens; October 13–19, 1989: 19–27.
7. Escobar J, Tuason V. Antidepression agents: A cross cultural study. Psychopharmacol Bull 1980; 16:49–52.
8. Jablensky A, Sartorious N, Emberg G, et al. Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization ten-country study. Psychol Med Monogr Suppl 1992: 20.

National Institute of Mental Health:

An Update on Women and Minorities in Research

Rick A. Martinez, M.D.

Abstract

Guidelines for the recruitment and retention of women and minorities as subjects in clinical trials have been developed at the National Institutes of Health. Training more women and minority investigators is also a National Institutes of Health priority. These individuals, like all other applicants, however, must be sophisticated consumers of the federal grants process. This article presents information about administration changes for entry-level research awards to help potential applicants understand the funding process.

Biomedical and behavioral clinical research supported by the National Institutes of Health (NIH) ultimately leads to changes in standards of care and health policy. Therefore, when claims are made about the efficacy and effectiveness of new interventions, it is important to understand these effects in women and members of minority groups. Congressional and scientific concerns about the lack of generalizing clinical research data have been addressed by legislation mandating administrative efforts to improve the quality of this work. These efforts are directed toward improving the recruitment of subjects in clinical trials and training women and minority investigators. This article begins with a review of certain changes in NIH policy that are designed to address this issue, mentions important information about streamlining National Institute of Mental Health (NIMH) grants, and offers pointers on planning a research career. Although the focus of this review deals with topics concerning women and minorities, the information on funding and developing a research career is applicable to all candidates for federal research awards, regardless of gender or race.

The Inclusion of Women and Minorities in Study Populations

The U.S. Congress addressed the historical underrepresentation of women and minorities in clinical treatment trials with special legislation in the National Institutes of Health Reauthorization Act of 1993, (*Federal Register* March 28, 1994; reprinted in *NIH Guide for Grants and Contracts,* March 18, 1994, vol. 23, no. 11). The legislation revised earlier guidelines to strengthen previous policies in place since 1990 and applies to all grants submitted after June 1, 1994. This legislation launched a flurry of activity at each of the 17 institutes to relieve problems of sample bias in clinical studies that receive federal funding. The Revitalization Act authorized the NIH Offices of Research on Women's Health and Minority Health to establish guidelines for peer-reviewed clinical studies. The new law requires each of the institutes to ensure that women and minorities are included in such research except in situations where the trials are inappropriate with respect to the health of the subjects or purpose of the research (for example, a prostate cancer trial would exclude women). In particular, phase III

From NIMH/Mental Disorders of the Aging Research Branch, Rockville, Maryland. Address reprint requests to Rick A. Martinez, M.D., Chief, Geriatric Psychiatry Research Program, NIMH/Mental Disorders of the Aging Research Branch, 5600 Fishers Lane, Room 18-105, Rockville, MD 20857.

clinical trials must include both groups in treatment studies. The law also prohibits the use of cost as a consideration for excluding members of each group from such trials.

Impact of the Reauthorization Act

Institutional review boards (IRBs, also known as study sections) have the primary responsibility for ensuring that these guidelines are implemented in their review of research applications. The institute realizes that it will take time for the field to understand and assimilate the guidelines, but the formal review process is actively engaged in evaluating grant applications under the new guidelines. To address questions about these guidelines, NIH staff have prepared a handbook called "NIH Outreach on the Inclusion of Women and Minorities in Biomedical and Behavioral Research" for review by potential candidates. The booklet contains general information about the guidelines as well as advice to researchers on how to enhance outreach efforts to improve recruitment and retention of women and minority subjects. This booklet can be obtained by writing the Office of Research on Women's Health at NIH, Building 10, Room 201, 9000 Rockville Pike, Bethesda, MD 20892 or by FAX (301) 402-1798.

National Institute of Mental Health

In fiscal year 1995 the bulk of the $630 million appropriated to NIMH was spent maintaining a national academic infrastructure designed to advance the treatment and prevention of mental illness. NIMH continues to provide funding for this research through three of its divisions: Neuroscience and Behavior (supports preclinical neurobiology, psychopharmacology, and cognitive studies), Biometry and Applied Science (mental health service provisions in other types of health settings such as primary care), and Clinical and Treatment Research (epidemiology, etiology, diagnosis, treatment, and prevention research of separate mental disorders).

Changes in NIMH Grants and Award Mechanisms

The Application Packet

The Public Health Service 398 form is required for all new, revised, competing continuation, and supplemental research submissions to NIH. This application packet was revised in May 1995 and is available on request from the Grants Information Office, Division of Research Grants, National Institutes of Health, 6701 Rockledge Drive, Room 1040, Base 7710, Bethesda, MD 20892-7710; phone (301) 435-0714. (You may e-mail requests and inquiries to asknih@odrockm1.od.nih.gov). Investigators should anticipate that NIH will eventually be able to accept electronic submission of grants; information about this development will be published in the *NIH Guide*.

Revision in Career Awards

NIH research award mechanisms for individual investigators are organized by career level, the type of previous training, and whether a mentor is required. Recent changes in these mechanisms to streamline the process were completed and announced in May 1995. Fourteen previous mechanisms have now been replaced by six new awards. For example, individuals with health-professional degrees who need support to obtain an intensive, supervised research experience may apply for the Mentored Clinical Scientists Development (KO8) Award. This award replaces four older mechanisms including the former Scientist Development Award for Clinicians (K20). Specific details about these and other changes can be obtained from the *NIH Guide* (April 28, 1995, vol. 24, no. 12).

NIMH Minority and Women's Supplements

In the fall of 1993, NIH announced an "emphasis" on the use of administrative supplements to attract underrepresented minorities into biomedical and behavioral research [see *NIH Guide* (November 26, 1993, vol. 22, no. 43)]. The supplemental funding, made available to a "parent grant," is designed for "junior" researchers (graduate and postgraduates) but includes high school and undergraduate minority individuals as well. The actual application process is initiated by the principle investigator who must submit the support materials and the request for supplemental funds on behalf of the minority individual. The application packet is submitted to the NIMH program staff, not the Division of Research Grants. NIH has also moved forward with efforts to increase the number of women researchers with modified versions of the administrative supplement. Principle investigators are advised to contact program staff before initiating this award mechanism.

Planning a Research Career

Formulating a Research Question

The competitive nature of applying for federal research grants and the myriad of award mechanisms available may at first seem intimidating, but candidates must consider several options in planning an academic career. In most cases, junior investigators are exposed to research occurs during their clinical research fellowship. After completing the fellowship the new faculty member must determine an area of research interest; collecting pilot data in the area of interest is critical. This process helps to focus the young investigator's questions, refine methodology, and generate hypotheses for systematic study. Regarding a research topic, some junior researchers spend a great deal of time anticipating which topics or ideas for research are the most "fundable" and assume that agency staff possess some secret knowledge about research ideas. The truth is, ideas for projects come from the field, not NIMH staff. Agency and institute priorities emerge from scientific perspectives that are crafted and developed in public forums such as scientific meetings, consensus conferences, and medical journals. This body of scientific information is the best source of information for generating hypotheses for investigation.

Finding Mentors

Finding a mentor is an important step in career planning. Senior faculty members, who have successfully competed for federal (and nonfederal) research money, are an invaluable resource to junior faculty members. Ideally, the counsel from an experienced mentor can help the researcher to organize an academic career plan with advice on scientific topics and tips on preparing a federal award application. NIMH program staff can also make useful contributions in terms of discussing ideas for possible studies, reviewing draft proposals, and providing general technical assistance around a particular research project or idea.

Choosing the Right Award Mechanism

Once potential researchers have decided to commit time to a research career, applicants can take advantage of entry-level NIH career awards to support their efforts. These grant awards provide significant portions of the investigators' salaries to relieve them of clinical responsibilities that can draw time away from research activities. The Mentored Clinical Scientist Award (KO8), mentioned earlier, provides salary and up to $50,000 for research development support for tuition, fees, supplies, and equipment. For the candidate with a research doctorate (and in some cases a health-professional doctorate) a comparable mentored experience is provided with the Mentored Research Scientist Development Award (KO1). After 3 to 5 years of support from the mentored awards, investigators will have had sufficient time to accumulate enough pilot data and research experience. They can then apply for an award as an independent investigator. NIH award mechanisms conceived for these newly independent investigators include the First Independent Research Support and Transition Award (R29) and the Research Project Grant (RO1) for more senior faculty.

Conclusion

NIH and NIMH, in particular, seek to address the underrepresentation of minorities and women in research and to increase their numbers in clinical trials and among the ranks of investigators who receive federal funding. Recent legislation is expected to remedy the situation, but the strategy to increase the representation of both groups among the ranks of researchers is more complex because planning an academic research career in medicine is no simple matter. The commitment in time, effort, and patience extract a great deal from individuals who choose to balance clinical responsibilities and generating pilot data with other important personal obligations. Getting started requires supportive department chairs and deans who can help launch an academic career. This is crucial because getting an application funded usually takes several submissions before it finally receives funding. As a result, maintaining this career requires that new faculty, regardless of gender or race, must become sophisticated consumers of NIH programs and federal (and nonfederal) funding mechanisms.

Sarcoidosis Mythology

Alvin S. Teirstein, M.D., Maria L. Padilla, M.D., Louis R. De Palo, M.D., and Gregory J. Schilero, M.D.

Abstract

Background. Sarcoidosis continues to be shrouded by anecdotal misinformation which has gained credence by repetition. These myths have been developing for the past 50 years and continue to accumulate, despite remedial data. Among the most egregious myths are that sarcoidosis is a disease of Blacks, that the chest radiograph is diagnostic of sarcoidosis, and has chronologic significance, that serum angiotensin-converting enzyme and bronchoalveolar lavage are diagnostic of sarcoidosis and serve as guides to therapy, that the Kveim-Siltzbach test is not a reliable diagnostic test for sarcoidosis, that sarcoidosis is difficult to diagnose, and that sarcoidosis is tuberculosis.

Methods and Results. The literature regarding these myths has been reviewed and supported by the experience with more than 10,000 patients with sarcoidosis who have been treated at Mount Sinai Medical Center, New York.

Conclusions. Sarcoidosis occurs with varying frequency among all races. The chest radiograph typical of sarcoidosis can be mimicked by other granulomatous and neoplastic diseases. The classic radiographic stages, from 0 to III, do not reflect the time course of sarcoidosis. The Kveim-Siltzbach test is an excellent diagnostic procedure. The diagnosis of sarcoidosis can be made relatively easily in most patients, but its etiology is still unknown.

Sarcoidosis Mythology

With the exception of infectious and inherited genetic disorders, a clearly delineated etiology is lacking for most diseases. Despite this important missing agent(s), clinicians have little difficulty recognizing and defining the vast array of human afflictions. For example, diabetes mellitus can be defined by the presence of hyperglycemia and rheumatoid arthritis can be identified by a typical clinical, pathologic, and serologic presentation. One would assume that the diagnosis of sarcoidosis as a disease presenting with noncaseating epithelioid granulomas in multiple organs, without identification of a known cause of granuloma formation, would offer clinicians similar confidence in dealing with this disease. Unfortunately, to this date, sarcoidosis remains a poorly understood entity of uncertain etiology, based on nonspecific pathology, which pursues a variable course and has unclear indications for treatment with a variety of unproven therapies.

Except for a brief spate of immunologic studies, very little scientific research has been directed at solving the unknown mechanisms that initiate the cascade of events from alveolitis to granuloma formation and eventual fibrosis in sarcoidosis. Instead, a large body of anecdotal information has accumulated which has served to obfuscate, rather than clarify, the clinical features of the disease. These data have given rise to an inordinate number of misconceptions which can be described as myths. Some of the most glaring examples of this mythology are:

1. Sarcoidosis is a disease of black Americans.
2. The chest radiograph is diagnostic of sarcoidosis.

This research was supported by The Catherine & Henry J. Gaisman Foundation, New York.

From The Mount Sinai Medical Center, Division of Pulmonary and Critical Care Medicine, New York, New York. Address reprint requests to Alvin S. Teirstein, M.D., Mount Sinai Medical Center, Division of Pulmonary and Critical Care Medicine, One Gustave L. Levy Place, Box 1232, New York, NY 10029.

3. Radiographic staging of chest radiographs has chronologic significance.
4. Serum angiotensin-converting enzyme determinations are diagnostic of sarcoidosis and serve as a guide to therapy.
5. Bronchoalveolar lavage leukocyte differential cell counts are diagnostic of sarcoidosis and serve as a guide to therapy.
6. Sarcoidosis is difficult to diagnose.
7. Sarcoidosis is tuberculosis.

Sarcoidosis Is a Disease of Black Americans

Beginning in 1961, several authors have reported ratios of sarcoidosis in Blacks compared with Whites as high as 20:1 (1–3). Studies emanating from major inner-city hospitals in Baltimore, Los Angeles, Philadelphia, and New York have supported this remarkable propensity of Blacks for the disease. However, knowledge of the history of sarcoidosis reveals that almost all original reports emanated from northern Europe, with the greatest incidence of disease occurring in Scandinavia, closely followed by England, France, and Germany, countries with relatively few black inhabitants (4). In New York, a comparative study of patients seen simultaneously in clinic and private offices of the same physicians (**Table 1**) revealed that of 104 clinic patients 57% were black and 31% were white compared with 110 more affluent patients seen in private offices where only 16% were black and 81% were white. Furthermore, a review of the Arthritis Clinic at the same institution revealed a preponderance of Blacks over Whites, similar to that seen in the sarcoidosis clinics (5). This suggests that reports of racial frequency of sarcoidosis, and many other diseases, gleaned from American urban hospital clinics reflect the racial composition of the clinic population rather than an accurate reflection of racial distribution of the disease in the United States. Currently, sarcoidosis is recognized as a major health problem in Japan where cardiac sarcoidosis is the most common cause of death in patients with the disease (6). Sarcoidosis has been reported in every continent and in almost every nation. Sarcoidosis is common in all races and in every geographic location; ethnicity is irrelevant in the diagnosis of this disease.

TABLE 1

Racial Distribution of Patients with Sarcoidosis Seen in Clinic and Private Offices vs Patients with Arthritis Seen in Clinic

	n	Black (%)	White (%)	Other (%)
Sarcoidosis				
Clinic	104	57	31	12
Private offices	110	16	81	3
Arthritis				
Clinic	121	58	28	14

At the Mount Sinai Medical Center, New York the ethnic distribution of sarcoidosis reflects the outpatient setting in which patients are seen. In clinics, the majority of patients are black. In private offices, most patients are white.

The Chest Radiograph Is Diagnostic of Sarcoidosis

Traditionally, chest radiographs of patients with sarcoidosis have been classified as stage 0, normal; stage I, bilateral hilar lymph node enlargement with clear lung fields; stage II, bilateral hilar lymph node enlargement with pulmonary infiltrations; stage III, pulmonary infiltrations with no lymph node enlargement (7). In 1973, Winterbauer et al. (8) wrote that in young, asymptomatic patients presenting with a chest radiograph exhibiting bilateral hilar adenopathy (stage I), sarcoidosis could be diagnosed without tissue confirmation. In Scandinavia and other countries, the co-existence of erythema nodosum with hilar lymph node enlargement, rightfully, is considered diagnostic of sarcoidosis. However, in the absence of erythema nodosum none of the radiographic presentations of sarcoidosis are diagnostic. Hodgkin's and non-Hodgkin's lymphoma, tuberculosis, histoplasmosis, coccidiomycosis, metastatic carcinomas and sarcomas, and acquired immunodeficiency syndrome (AIDS) may result in chest radiographs that are indistinguishable from sarcoidosis (9) (**Fig. 1**). Although Winterbauer's certainty of diagnosis of sarcoidosis by chest radiograph alone may be correct in the majority of young patients, when an error is made, diagnosis of a life-threatening disease may be delayed. If tissue proof of sarcoidosis were difficult to obtain, it might be acceptable to base the diagnosis on the radiographic patterns alone. However, as will be discussed later, tissue diagnosis in sarcoidosis usually is easy to obtain.

Radiographic Staging of Sarcoidosis Has Chronologic Significance

A system that classifies "stages" from 0 to III implies disease chronology. Although some patients have progressed from one stage to another, such progression cannot be demonstrated in most patients. Israel et al. (10) have shown that, in 25% of patients presenting with stage I radiographs, hilar lymph node enlargement persists for years. Thus, faced with a patient who exhibits a stage I radiograph, the clinician does not know

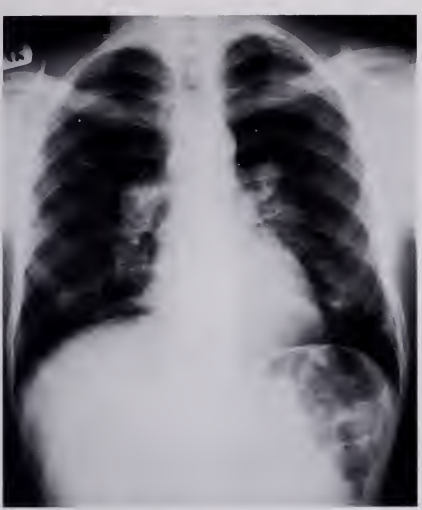


Fig. 1. Chest radiograph of a 29-year-old man with AIDS. The radiograph shows bilateral hilar lymph node enlargement and pulmonary nodules indistinguishable from stage II sarcoidosis. Mediastinoscopic lymph node biopsy and repeated sputum cultures revealed Mycobacterium avium-intracellulare.

if this is acute or chronic disease, unless a previous chest radiograph is available. Because sarcoidosis occurs predominantly in the young age group, old chest radiographs usually do not exist. Other patients, however, present with acute onset of dyspnea and radiographs revealing bilateral pulmonary infiltrations (stage III) with an apparent illness of only a few weeks or months duration. Thus, stage I radiographic pattern may occur in patients with chronic sarcoidosis and stage III in those with acute disease. Finally, the radiograph staging system is faulted by large reader variability. Qualified radiologists and pulmonologists will frequently differ when the radiologic patterns border on the interfaces of the four stages. Only stage IV radiographs (not included in the original staging schema), which show evidence of pulmonary fibrosis, including bullous transformation, cephalad retraction of the hilar areas, and tenting of the diaphragm and aspergillomas, have chronologic significance. Stage IV radiographs always indicate chronic disease.

Serum Angiotensin-Converting Enzyme Determination Is Diagnostic of Sarcoidosis and Serves as a Guide to Therapy

In a series of papers beginning in 1975, Liebermann and others (11–13) reported that an elevated serum angiotensin-converting enzyme (SACE) level was diagnostic of sarcoidosis and reflected disease activity, serving as a reliable guide to therapy. However, subsequent studies have demonstrated that elevated SACE levels occur with varying frequency in patients with sarcoidosis, ranging from 40 to 80%, with most studies reporting only 50% of patients with elevated SACE levels (14–16). Furthermore, elevated SACE levels have been reported in a variety of nonsarcoid diseases, many of which may mimic sarcoidosis (17).

Although it was true that an elevated SACE level in patients with proven sarcoidosis indicates the presence of active disease, therapy initiated solely because of an increased SACE level has not been shown to alter the eventual outcome for patients. Even more troublesome is the absence of increased SACE levels in many patients with active disease as determined by the clinical presentation, supported by the chest radiograph and pulmonary function tests. For example, the most clinically active presentation of sarcoidosis is the presence of erythema nodosom (Löfgren's syndrome) (18). In a study of 31 patients with Löfgren's syndrome, only 55% had elevated SACE levels (19). Therefore, SACE is not diagnostic of sarcoidosis and is an unreliable guide to therapy.

Bronchoalveolar Lavage Leukocyte Differential Is Diagnostic of Sarcoidosis and Serves as a Guide to Therapy

Bronchoalveolar lavage (BAL) has proven to be a revolutionary technique in the diagnosis of diffuse lung diseases caused by specific recognizable etiologic agents (20). Diffuse lung diseases involving the alveoli and/or pulmonary interstitium which present with nonspecific bilateral radiographic infiltrations are an everyday occurrence. Diagnostic BAL is routinely performed in these patients. The procedure has removed much of the diagnostic uncertainty, especially in patients with opportunistic lung infections (21). The use of BAL to sample bronchoalveolar cellular and biochemical constituents has been a major impetus for research into the cellular and immunologic events which occur in all interstitial lung diseases and asthma. For decades, bronchoscopic biopsy has been the accepted standard for the diagnosis of sarcoidosis in many patients (22,23). In 1981, Hunnighake and Crystal (24) proposed that a CD_4 lymphocytic BAL leukocyte differential count without biopsy was diagnostic of sarcoidosis and BAL with increased polymorphonuclear cells implied a diagnosis of idiopathic pulmonary fibrosis (IPF). Although it is true that a lymphocytic BAL cell differential in patients with a typical clinical and radiologic presentation of sarcoidosis is common, subsequent studies have revealed that BAL lymphocytosis may be present in a variety of diffuse lung diseases, many of which exhibit chest radiographic patterns indistinguishable from sarcoidosis (20). **Table 2** shows a comparison of the results of BAL cell count differentials in 46 patients with sarcoidosis and 55 patients with IPF. A bare majority of the patients with IPF (28 of 55) had increased BAL polymorphonuclear leucocytes. Approximately three fourths of the patients with sarcoidosis had BAL lymphocytosis. Had BAL been relied on as a diagnostic test, one fourth of the patients with sarcoidosis would not have received a diagnosis, and the 12 patients with IPF and lymphocytosis would have received a misdiagnosis of sarcoidosis (9). Subsequently, it was suggested that BAL lymphocytosis indicated the presence of active sarcoidosis and was a valuable guide to therapy (25). However, studies using clinical indices of activity including symptoms, worsening chest radiographs and pulmonary function studies, the presence of extrathoracic disease, and response to steroid therapy, revealed that "active" sarcoidosis exists in many patients without demonstrable BAL-accessed lymphocytosis (9). **Table 3** shows the results of BAL in 46 patients with sarcoidosis. Twenty-seven patients were deemed to have active disease and 19 had inactive, using clinical, radiographic, and physiologic parameters. Six of the 27 active patients did not have BAL lymphocytosis and 11 of the 19 inactive patients exhibited BAL lymphocytosis. Thus, BAL lymphocytosis in this study was a poor guide to therapy. The distribution of lymphocytes in the lungs of patients with sarcoidosis is often heterogeneous, leading to sampling errors; BAL lymphocytosis may persist even when the disease has abated clinically and radiographically (26). It must be concluded that BAL without lung biopsy is not diagnostic of sarcoidosis and is an uncertain and impractical guide to therapy.

TABLE 2
BAL Leukocyte Differential in 101 Patients with IPF and Sarcoidosis

	n	POLYS > 15%	LYMPHS > 20%	Normal
IPF	55	28	12	15
Sarcoid	46	2	33	11

Most patients with IPF had increased polymorphonuclear cells (POLYS) and most with sarcoidosis had BAL lymphocytosis (LYMPHS). However, if BAL differential cell counts were used for diagnosis, almost one half the patients with IPF and more than one fourth of the patients with sarcoidosis would not have received the correct diagnosis.

TABLE 3
Results of BAL Leukocyte Differential Counts in 46 Patients with Sarcoidosis, by Activity

	n	POLYS >15%	LYMPHS > 28%	Normal
Active	27	1	21	5
Inactive	19	1	11	6

Using clinical, radiographic, and physiologic data, 46 patients with sarcoidosis were divided into active (27) and inactive (19). BAL leukocyte differential counts misclassified more than 20% of the active patients and almost 60% of the inactive patients. POLYS = polymorphonuclear cells; LYMPHS = BAL lymphocytosis.

The Kveim-Siltzbach Test Is Not a Reliable Diagnostic Test for Sarcoidosis

In 1941, Kveim reported the occurrence of sarcoid-like granulomas in 13 patients with sarcoidosis at the injection site of a suspension derived from a lymph node taken from a single patient with the disease (27). These findings were confirmed by Williams and Nickerson (28), Chase (29), and Siltzbach (30). In an international trial of more than 3,000 patients, Siltzbach (31) firmly established the diagnostic validity of the Kveim test with which his name is now coupled. Subsequently, a remarkable number of authors reported false-positive Kveim-Siltzbach (KS) tests in a variety of illnesses including tuberculosis (32), infectious mononucleosis, lymphatic leukemia (33), Crohn's disease (34), and lupus erythematosis (35). When positive KS tests were obtained in healthy subjects, Izumi et al. (36) were able to trace the spate of false-positive tests to a new batch of commercially distributed suspensions, which fortunately have been withdrawn from the market. At the Mount Sinai Medical Center, New York, the KS test has been in constant use since 1948. In 1988, 78% true positives in patients with sarcoidosis and only one false positive in patients without sarcoidosis was reported from that center (37). The KS test, currently used throughout Europe and Japan, has proven to be a valuable, reliable diagnostic procedure. Indeed, it is the only test that allows a specific diagnosis of sarcoidosis, rather than a pathology report of a tissue biopsy which usually proclaims the presence of "noncaseating epithelioid granulomas without acid fast bacilli or fungi, consistent with sarcoidosis."

The Diagnosis of Sarcoidosis Is Difficult

In most patients, the diagnosis of sarcoidosis is easy. Biopsies of cutaneous granulomatous lesions (not erythema nodosum), peripheral lymph nodes, granulomatous conjunctiva, and lacrimal glands can be done under local anesthesia as an outpatient procedure. If these sites are not available, bronchoscopic lung biopsy, scalene lymph node biopsy, and mediastinoscopy offer diagnostic sites. Neurologic and cardiac sarcoidosis, in the absence of other more accessible lesions, may present diagnostic difficulties. Endomyocardial biopsies in patients with sarcoidosis reveal granulomas in less than 50% of patients (38). Brain biopsies are rarely required. In these patients, a computed tomographic scan of the chest and/or ^{67}Ga scan may uncover other more accessible biopsy sites not apparent by physical examination or routine radiographs. For those physicians with access to validated KS suspensions, none of the more invasive biopsy procedures are necessary in up to 80% of patients. In an international study of more than 200 centers, approximately half of which used KS tests, the order of preference following the skin test was bronchoscopy, scalene lymph node biopsy, and mediastinoscopy (39). A simple diagnostic algorithm is shown in Fig. 2.

Sarcoidosis Is Tuberculosis

From the time of the earliest reports, sarcoidosis was thought to be an aberrant form of tuberculosis. The presence of epithelioid granulomas in tissue, pulmonary and intrathoracic lymph node involvement, and dissemination throughout multiple organs suggested a tubercular etiology in an era when tuberculosis was very common. However, despite generations of careful study, the search for mycobacterium tuberculosis has been fruitless. Indeed, the current accepted criteria for diagnosis of sarcoidosis mandate the absence of acid fast bacilli in biopsy specimens. Clinical differences distinguish the two diseases. Disseminated cutaneous granulomas, "punched out" lesions in the phalanges, and salivary gland and lacrimal gland involvement are common in sarcoidosis and absent in tuberculosis. Therapeutic trials with antituberculosis therapy have not been effective in sarcoidosis and sarcoidosis granulomas are suppressed by adrenocortical steroids in contrast to tuberculosis.

Despite the weight of evidence against mycobacterium tuberculosis as the etiologic agent in sarcoidosis, a small, but growing number of studies have suggested the presence of mycobacteria in the blood and tissues of patients with sarcoid-

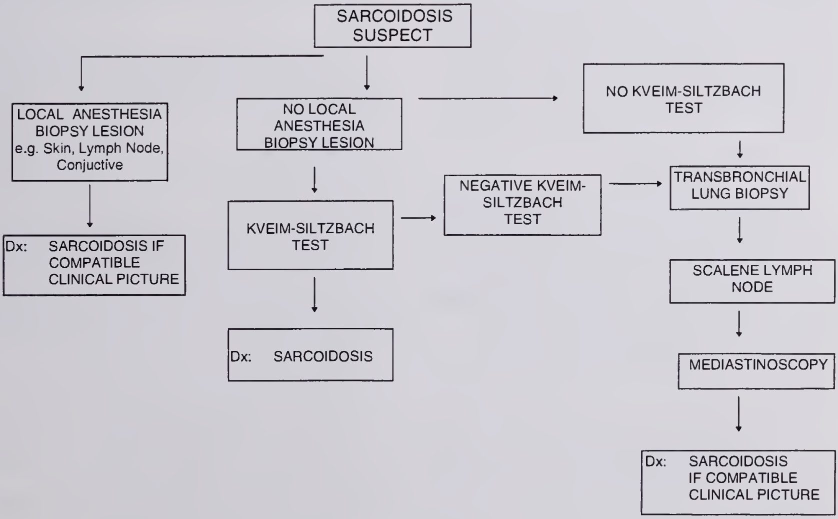

SARCOIDOSIS SUSPECT
LOCAL ANESTHESIA BIOPSY LESION e.g. Skin, Lymph Node, Conjuctive
Dx: SARCOIDOSIS IF COMPATIBLE CLINICAL PICTURE
NO LOCAL ANESTHESIA BIOPSY LESION
KVEIM-SILTZBACH TEST
Dx: SARCOIDOSIS
NEGATIVE KVEIM-SILTZBACH TEST
NO KVEIM-SILTZBACH TEST
TRANSBRONCHIAL LUNG BIOPSY
SCALENE LYMPH NODE
MEDIASTINOSCOPY
Dx: SARCOIDOSIS IF COMPATIBLE CLINICAL PICTURE

Fig. 2. Algorithm for the diagnosis of sarcoidosis.

osis. With the recent introduction of polymerase chain reaction (PCR) techniques, there has been a renewed interest in and growing number of conflicting reports concerning the presence of mycobacteria. In 1964, Mankiewicz (40) isolated mycobacteriophages in stool and tissues of patients with sarcoidosis. Serum mycobacterial antibodies were detected in patients with sarcoidosis by Chapman and Speight (41). Fluorescent microscopy of sarcoidal lymph node biopsies reportedly demonstrated mycobacteria (42). Judge and Mattman (43) have been growing cell wall-deficient mycobacteria from the blood of patients with sarcoidosis for the past 20 years, work now expanded by Almenoff et al. (44). To date, attempts to identify the organisms have failed. Using PCR, a rapidly growing number of reports have maintained that mycobacterial DNA can be found in the tissues and blood of patients with sarcoidosis (45–47). However, an equal number of PCR studies have refuted these claims (48–50). In a review of these conflicting reports, Mangiapan and Hance (51) emphasized the difficulty in applying current PCR techniques in studies of the etiology of infectious diseases and proposed that sarcoidosis was a paucibacillary state that requires techniques more specific than those available today in most laboratories.

With the lung as the portal of entry, and dissemination of granulomas throughout the body, most observers agree that sarcoidosis is probably caused by an infectious agent invading a susceptible host. It is unlikely that the offending agent(s) is mycobacterium tuberculosis, but it is predictable that the etiologic infection is mycobacterial-like. Identification of the organism is imminent.

Summary

Despite years of accumulating epidemiologic and clinical data, knowledge of sarcoidosis continues to be clouded by misinformation. None of these myths are supported by data and only serve to confuse the clinician in pursuit of the diagnosis and treatment. One must conclude that:

1. Sarcoidosis is *not* a disease of blacks.
2. The chest radiograph is *not* diagnostic of sarcoidosis.
3. The chest radiograph *does not* reflect the chronology of sarcoidosis.
4. SACE determinations and BAL are *not* diagnostic of sarcoidosis and *do not* serve as a guide to therapy.
5. The KS test *is* an excellent diagnostic procedure for sarcoidosis.
6. Sarcoidosis is *not* difficult to diagnose.
7. Sarcoidosis is *not* tuberculosis.

Acknowledgment

The assistance of Jessica Vélez-Benito is greatly appreciated.

References

1. Couch JW. Sarcoidosis in the United States Army, 1952 through 1956. Am Rev Respir Dis 1961; 84(Pt II):103–108.
2. Gundelfinger BF, Britten SA. Sarcoidosis in the United States Navy. Am Rev Respir Dis 1961; 84(Pt II):109–115.
3. Sartwell PE, Edwards LB. Epidemiology of sarcoidosis. Am J Epidemiol 1974; 94:250–257.
4. Teirstein AS, Lesser M. Worldwide distribution and epidemiology of sarcoidosis. In: Fanburg BL, editor. Sarcoidosis and other granulomatous disorders of the lung. New York: Marcel-Dekker, Inc., 1983: 101–134.
5. Teirstein AS, Siltzbach LE, Gabelman M, Lesser M. Socioeconomic factors influencing epidemiologic data in sarcoidosis among mixed ethnic urban population. In: Chretien J, Marsac J, Satiel JC, editors. Sarcoidosis and other granulomatous disorders. Oxford: Pergamon Press, 1981: 234–236.
6. Tachibana T, Kazaro I, Takemura T. Study on the cause of death in patients with sarcoidosis in Japan. Sarcoidosis 1992; 9:307–308.
7. Siltzbach LE. Sarcoidosis: clinical features and management. Med Clin North Am 1967; 51:483–502.
8. Winterbauer RH, Belie M, Moores KD. A clinical interpretation of bilateral hilar adenopathy. Ann Intern Med 1973; 78:65–71.
9. Teirstein AS. Sarcoidosis, a ship in search of a sail. In: James G, editor. Sarcoidosis and other granulomatous disorders. New York: Marcel Dekker, 1994; 73:745–761.
10. Israel HL, Sperber M, Steiner RM. Course of chronic hilar sarcoidosis in relation to markers of granulomatosis activity. Invest Radiol 1983; 18:1–5.
11. Lieberman J. Elevation of serum angiotensin-converting enzyme (ACE) levels in sarcoidosis. Am J Med 1975; 59:365–372.
12. Lieberman J, Nosal A, Scheissner CA, et al. Serum angiotensin-converting enzyme for diagnosis and therapeutic evaluation of sarcoidosis. Am Rev Respir Dis 1979; 120:329–355.
13. Romer FK. The level of angiotensin-converting enzyme as an indicator of 2 year prognosis in untreated pulmonary sarcoidosis without erythema nodosum. Acta Med Scand 1982; 211:293–295.
14. Gronhagen-Riska C. Angiotensin-converting enzyme: activity and correlation with serum lysosome in sarcoidosis, other chest or lymph node diseases and healthy persons. Scand J Respir Dis 1979; 60:83–93.
15. Romer FK. Clinical and biochemical aspects of sarcoidosis with special reference to angiotensin-converting enzyme (ACE). Acta Med Scand 1984; 69(Suppl):1–96.
16. De Remee RA, Rohrboch MS. Normal serum angiotensin converting enzyme activity in patients with newly diagnosed sarcoidosis. Chest 1984; 85:45–48.
17. Hollinger MA. Serum angiotensin-converting enzyme. Status report on its diagnostic significance in pulmonary disease. Chest 1983; 63:589–590.
18. Löfgren S, Lumlbäck H. The bilateral hilar lymph node

syndrome: A study of the relation to age and sex in 212 cases. Acta Med Scand 1952; 142:259–264.

19. Teirstein AS, Krumholtz S. Assessment of serum angiotensin-converting enzyme as a marker of activity in sarcoidosis: a study of 31 patients with erythema nodosum. Mt Sinai J Med 1987; 54:144–146.
20. Turner-Warwick MA, Harlem PL. Clinical application of bronchoalveolar lavage: an interim view. Br J Dis Chest 1986; 88:105–121.
21. Broaddus C, Dake MD, Stubbage MS, et al. Bronchoalveolar lavage and transbronchial biopsy for the diagnosis of pulmonary infiltrations in the acquired immunodeficiency syndrome. Ann Int Med 1985; 102:747–752.
22. Koerner SK, Sakowitz RI, Appelman, et al. Transbronchial lung biopsy for the diagnosis of sarcoidosis. N Engl J Med 1975; 293:268–270.
23. Teirstein AS, Chuang M, Miller A, Siltzbach LS. Flexible bronchoscope biopsy of lung and bronchial wall in intrathoracic sarcoidosis. Ann NY Acad Sci 1976; 278: 522–527.
24. Hunninghake GW, Crystal RG. Pulmonary sarcoidosis: a disorder mediated by excess helper T-lymphocyte activity at sites of disease activity. N Engl J Med 1981; 135: 250–263.
25. Klech H. Clinical risk assessment in sarcoidosis. In: Grassi C, Rizzato G, Pozzi E, editors. Sarcoidosis and other granulomatous disorders. Amsterdam: Excerpta Medica, 1988: 461–482.
26. Peterson MW, Nugent KM, Joller H, et al. Uniformity of bronchoalveolar lavage in patients with sarcoidosis. Am Rev Respir Dis 1988; 137:79–84.
27. Kveim A. En ng og spesifik kuloss-reackjon ved Beock's Sarcoid. Nord Med 1941; 9:169–172.
28. Williams RH, Nickerson DA. Skin reactions in sarcoidosis. Proc Soc Exp Biol Med 1935; 33:403–445.
29. Chase MW. Preparation and standardization of Kveim testing antigen. Am Rev Respir Dis 1961; 84(Pt 11):86–88.
30. Siltzbach LE. The Kveim test in sarcoidosis: a study of 750 patients. J Am Med Assoc 1961; 178:476–482.
31. Siltzbach LE. The international Kveim test study: 1960–1966. In: Turiaf T, Chabot J, editors. Proceedings of the IV International Conference on Sarcoidosis. Paris: Masson, 1900: 201–213.
32. Israel HL, Sones M. The diagnosis of sarcoidosis with special reference to Kveim reaction. Ann Intern Med 1955; 43:1268–1282.
33. Israel HL, Goldstein RA. Relation of Kveim antigen reaction to lymphadenopathy: a study of sarcoidosis and other diseases. N Engl J Med 1971; 284:345–349.
34. Karlish AJ, Cox EV, Hampson F, et al. Kveim test in Crohn's disease. Lancet 1970; 2:977–978.
35. Bringle C. Relationship between sarcoidosis and collagen diseases: preliminary report. Reports of the European Symposium on Sarcoidosis. Sarcoidosie 1972: 34–35.
36. Izumi T, Kobara Y, Morioka S, et al. False positive reaction in the Kveim test using the CSL Kveim material. In: Ivai K, Hoso Y, editors. Proceedings of the VI International Conference on Sarcoidosis. Japan: University of Tokyo Press, 1974: 77–78.
37. Teirstein AS, Brown LK. The Kveim test in 1987. In: Grassi C, Rizzato G, Pozzi E, editors. Sarcoidosis and other granulomatous disorders. Amsterdam: Elsevier, 1988: 7–18.
38. Johns CJ, Paz H, Kasper EK, Baugham K. Myocardial sarcoidosis: cause and management. Sarcoidosis 1992; 9:231–236.
39. Teirstein AS, Siltzbach LE, Dorph D. Report of international questionnaire regarding diagnostic procedures in sarcoidosis: the impact of fiberoptic bronchoscopy. In: Williams WJ, Davies BH, editors. Sarcoidosis. Cardiff: Alpha Omega Press, 1980: 230–237.
40. Mankiewicz E. The relationship of sarcoidosis to anonymous bacteria. Acta Med Scand 1964: (Suppl 425):68–73.
41. Chapman JS, Speight M. Further studies of mycobacterial antibodies in the sera of sarcoidosis patients. Acta Med Scand 1964; (Suppl 425):61–67.
42. Richter J, Bartak F, Halova R. Detection of mycobacteria by fluorescent microscopy in sarcoidosis; Prague 1971; Universita Karlova, 1971: 83–84.
43. Judge MS, Mattman LH. Cell wall deficient mycobacteria in tuberculosis, sarcoidosis and leprosy. In: Domingue, GJ ed. Cell Wall Deficient Bacteria. Reading, MA: Addison-Wesley: 1982: chapter 10.
44. Almenoff PL, Johnson A, Lesser M, Mattman LH. Growth of acid fast L forms from the blood of patients with sarcoidosis. Thorax. In press.
45. Saboor SA, Johnson NM, McFadden JJ. Detection of mycobacterial DNA in sarcoidosis and tuberculosis with polymerase chain reaction. Lancet 1992; 339:1012–1015.
46. Fiddler HM, Rook GA, Johnson NM, McFadden JJ. Mycobacterium tuberculosis DNA in tissue affected by sarcoidosis. Br Med J 1993; 306:546–549.
47. Mitchell IC, Turk JL, Mitchell DN. Detection of mycobacterial RNA in sarcoidosis with liquid-phase hybridization. Lancet 1992; 339:1015–1017.
48. Gerdes J, Richter E, Rüsch-Gerdes S, et al. Mycobacterial nucleic acids in sarcoid lesions. Lancet 1992; 339:1536–1537.
49. Thakker B, Black M, Foulis AK. Mycobacterial nucleic acids in sarcoid lesions. Lancet 1992; 339:1537.
50. Bocart D, Lecossier D, DeLassence A, et al. A search for mycobacterial DNA in granulomatous tissues from patients with sarcoidosis using polymerase chain reaction. Am Rev Respir Dis 1992; 145:1142–1148.
51. Mangiapan G, Hance AJ. Mycobacteria and sarcoidosis: an overview with summary of recent molecular biologic data. Sarcoidosis 1995; 12:20–37.

Fifty Years of Experience with Antithrombotic Therapy in Cardiac Disease

A 1996 Approach Based on Pathogenesis and Risk

RONALD S. FREUDENBERGER, M.D. AND VALENTIN FUSTER, M.D., PH.D.

SINCE THE DISCOVERY OF HEPARIN by McLean in 1916 (1) and the initial synthesis of acetylsalicylic acid by Hoffman in 1899 (2) and warfarin in the 1940s (3), the use of antiplatelet and antithrombotic agents in the treatment and prevention of cardiac disease has become widespread. In 1959, the first trial of thrombolytic agents for the treatment of acute myocardial infarction and other thromboembolic disorders in humans was conducted (4). In the 1960s and 1970s studies demonstrated the effectiveness of oral anticoagulants for the early treatment of acute myocardial infarction (5, 6). In the 1980s the efficacy of aspirin in acute coronary syndromes became convincing. Since these initial discoveries, it has become evident that thrombosis in cardiovascular disease is a central mechanism responsible for much morbidity and mortality. Thus, a sound understanding of the pathogenesis of the thrombotic processes in cardiovascular disease and the relative risks of thrombosis and thromboembolism is required to develop a rational approach to antithrombotic therapy. Based on the experience of the past 50 years, this review will provide guidelines for antithrombotic therapy in the various cardiac syndromes from the perspective of pathogenesis and thrombotic and thromboembolic risk. This information is summarized in Table 1.

Pathogenesis of Thrombosis and Thromboembolism: Platelets and/or Fibrin

Coronary Arteries

Vascular injury and thrombus formation are central events in the initiation and progression of the atherosclerotic process leading to the acute coronary syndromes. When there is superficial endothelial injury, the platelets form a monolayer to cover the exposed subendothelium. When injury is more severe, such as plaque rupture during the acute coronary syndromes or immediately after angioplasty, exposure of the vessel wall substrate, such as the lipid pool, appears to lead to fibrin formation and subsequent platelet activation and aggregation by exposure of the platelet membrane IIb/IIIa receptors (Fig. 1). This results in platelet aggregation which in turn leads to further activation of the intrinsic and extrinsic coagulation systems (7). Specifically, at the site of plaque damage, the extrinsic pathway appears to be activated through the release of tissue factor in part from the macrophages that are present in the arterial wall (7), and it is the generation of thrombin which leads to fibrin formation and eventually contributes to platelet activation.

Acute Coronary Syndromes. In many patients with unstable angina, plaque rupture leads to thrombus formation and transient occlusion of the coronary artery, usually lasting only 10 to 20 minutes, followed by reperfusion of the vessel (7).

Assistant Professor of Medicine, Director, Cardiac Transplantation (RSM), The University of Maryland School of Medicine, Baltimore, Maryland and Professor of Medicine, Director, The Cardiovascular Institute, Dean for Academic Affairs (VF), The Mount Sinai Medical Center, New York, New York. Address correspondence to Valentin Fuster, M.D., Ph.D., Director, The Cardiovascular Institute, Mount Sinai Medical Center, Box 1030, 1 Gustave Levy Place, New York, NY 10029.

TABLE 1

Thromboembolism in Cardiovascular Disease Based on Pathogenesis and Risk: Use of Anticoagulants and Aspirin[a]

	Thromboembolic Risk		
Pathogenesis	High (>6% yr)	Medium (2–6% yr)	Low (<2% yr)
Arterial	ACS, PTCA	Stable CAD	Primary prevention
Platelets and fibrin (PI and/or A/C)	ASA + A/C	ASA or A/C	PI ?
Chambers	A-fib–emboli	A-fib with or without valvular disease	A-fib–lone chronic LV aneurysm
	A-fib–mitral stenosis	Anterior MI–early dilated cardiomyopathy	
Fibrin (A/C)	INR 2.5–3.5	INR 2.0–3.0	No therapy
Prostheses	Old mechanical	Recent mechanical INR 2.5–3.5	Bioprosthesis–normal sinus rhythm
Fibrin more than platelets (A/C > PI)	Previous emboli Extensive atherosclerosis INR 3.0–4.5 or INR 2.5–3.5 + PI	Bioprosthesis–A-fib INR 2.0–3.0	No therapy

[a] PI = platelet inhibitor (i.e., aspirin); ASA = aspirin; A/C = anticoagulants; ACS = acute coronary syndrome; PTCA = percutaneous transluminal coronary angioplasty; CAD = coronary artery disease; A-fib = atrial fibrillation; LV = left ventricle; INR = International Normalized Ratio; MI = myocardial infarction.

Modified from Stein B, Fuster V, Halperin JL, Chesebro JH. Circulation 1989; 80:1501.

In non-Q-wave myocardial infarction the plaque damage is probably more severe and occlusion of the vessel more persistent, often lasting as long as 1 hour. In Q-wave infarction there is persistent occlusion of the coronary artery, which leads to cessation of blood flow for more than 1 hour, resulting in transmural cellular necrosis.

Intermittent occlusion occurs because the occlusive thrombus may undergo spontaneous lysis. However, the intra-arterial surface (i.e., the ruptured plaque and the thrombus) remains very thrombogenic, thus creating a high risk of rethrombosis. It appears that the surface of the remaining thrombus exposes thrombin which activates platelets and the coagulation cascade. Thus, activation of both the coagulation system and

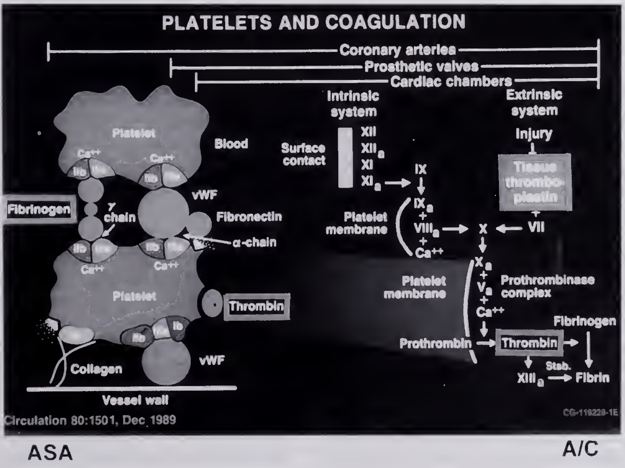

PLATELETS AND COAGULATION
Coronary arteries
Prosthetic valves
Cardiac chambers
Intrinsic system
Extrinsic system
Platelet
Blood
Ca++
IIb
IIIa
vWF
Fibrinogen
γ chain
Fibronectin
α-chain
Platelet
Thrombin
Collagen
Vessel wall
Surface contact
XII
XIIa
XI
XIa
IX
IXa
Platelet membrane
VIIIa
X
VII
Injury
Tissue thromboplastin
Xa
Va
Platelet membrane
Prothrombinase complex
Prothrombin
Thrombin
Fibrinogen
Stab.
XIIIa
Fibrin
Circulation 80:1501, Dec 1989
CG-119228-1E
ASA
A/C

Fig. 1. Coagulation scheme initiated by tissue injury or mechanical device.

platelets plays a pivotal role in thrombus formation in the acute coronary syndromes (7).

Coronary Intervention. Percutaneous transluminal coronary angioplasty (PTCA) and other catheter-based interventional techniques are associated with denudation of the endothelium and deep vessel injury. This results in exposure of thrombogenic elements in the atherosclerotic plaque and the vessel wall, which predisposes this area to fibrin formation and platelet deposition leading to intravascular thrombosis (8). Therefore, therapy to prevent acute occlusion after coronary intervention should be directed against both platelet activation and the coagulation cascade.

Saphenous Vein Bypass Grafting. Disease of the saphenous vein graft is the most important factor leading to recurrent ischemia after coronary bypass surgery. Occlusion rates are 8 to 18% per distal anastomosis at 1 month postoperatively and 16 to 26% at 12 months (9). Vein graft disease can be divided into three phases: an early postoperative phase (within 1 month after surgery) related to thrombotic occlusion in which platelet activation and fibrin formation are implicated; an intermediate phase (within the first postoperative year) characterized by intimal hyperplasia resulting in a form of accelerated atherosclerosis that may have a superimposed thrombotic tendency; and a late phase (after the first postoperative year) characterized by graft atherosclerosis similar to that of the native coronary arteries (9). Therefore, the predominant pathogenetic mechanism in saphenous vein bypass grafting is related to both platelet activation and fibrin generation, particularly in the early postoperative phase.

Progression of Chronic Coronary Disease. One might expect atherosclerotic lesions to expand in a linear fashion. However, angiographic studies have shown this not to be the case. Many high-grade lesions often appear in segments of the coronary artery that were previously normal (10). This erratic and unpredictable growth may be caused by plaque disruption or fissuring and intracoronary mural thrombosis (7). Thus, mural thrombus forms by the same mechanism as thrombus in the acute coronary syndromes; but, in most patients, the thrombus is not occlusive. The mural thrombus then undergoes fibrous organization and contributes, often asymptomatically, to the progression of the disease.

Cardiac Chambers

We will relate this brief discussion of intracavitary thrombus formation to that of the left ventricle and left atrium.

Left Ventricle. The development of intracavitary or mural thrombi in patients with acute myocardial infarction, left ventricular aneurysm, and dilated cardiomyopathy is well described. The pathogenesis of intracavitary thrombosis is related to the classical description, by Virchow (11), who defined three precipitating factors: endothelial injury, stasis, and hypercoagulability. Stasis and hypercoagulability contribute significantly to fibrin deposition and activation of the coagulation system (Fig. 1). Once a thrombus forms, the clinical significance is mainly related to its potential for embolism.

Endocardial Injury. After an acute myocardial infarction, leukocyte infiltration separates endothelial cells from their basal lamina. This results in exposure of the subendothelial tissue to intracavitary blood, which leads to activation of platelets and the coagulation system and may trigger thrombus formation.

Blood Stasis. Blood stasis in areas of akinesis or dyskinesis as well as in a diffusely or segmentally dilated ventricle with or without cardiac failure, is an important factor in predisposition to thrombus formation. The three most classical clinical conditions predisposed to blood stasis are dilated cardiomyopathy, anterior myocardial infarction, and left ventricular aneurysm. Stasis triggers activation of the coagulation system, leads to fibrin formation, and accounts for the predominant pathogenetic mechanism in the development of intracavitary thrombus formation.

Hypercoagulable State. During an acute myocardial infarction there is a greater tendency toward systemic coagulation. This is based on observations of increased deep venous thrombosis and left ventricular thrombus formation and on experimental evidence of systemic elevations of serum fibrinogen levels.

Dynamic Forces of the Circulation. Thus, although the combination of the three components of Virchows' triad predisposes to thrombus formation, formation of intracavitary thrombi does not necessarily lead to embolization. Dynamic forces of the chambers determine the propensity toward embolization. For example, in the case of left ventricular aneurysm, the isolation from dynamic circulatory forces may protect from systemic embolization; this may partly explain the relatively low incidence of thromboembolism in segmental ischemic cardiomyopathy. In contrast, in dilated cardiomyopathy, a diffuse process, the mural thrombus is not isolated from the circulation, and therefore may be at higher risk of embolization. Therefore, when stratifying risk, one must consider not only the predisposition to thrombus formation but also the risk of subsequent embolization.

Left Atrium. The pathophysiology of thrombus formation in the left atrium is similar to that described above in which stasis is the predominant mechanism. In nonvalvular atrial fibrillation, stasis leads to fibrin generation and potentially to thrombus formation (Fig. 1). Valvular heart disease, particularly mitral stenosis, also predisposes to fibrin generation by increasing stasis in the atria. When atrial fibrillation occurs with mitral stenosis, the risk of thrombus formation in the left atrium becomes high.

Thus, in the cardiac chambers, thrombus formation occurs because of stasis, endocardial injury, and hypercoagulability. The predominant pathophysiological process is related to generation of fibrin.

Prosthetic Valves

Thrombus formation in mechanical prosthetic heart valves and bioprosthetic valves will be described in the next section.

Mechanical Prostheses. Almost immediately after implantation of a mechanical prosthesis, platelet deposition begins, particularly on the endocardium-suture-prosthesis interfaces and on damaged perivascular tissue. Platelet deposition leads to activation of factor XII and initiation of the coagulation cascade (12). Eventually, stasis and abnormal hemodynamics promote fibrin generation which predominates over the platelet component of this response (Fig. 1). This is particularly true when there is left atrial enlargement which allows for a greater degree of stasis. In recent years, efforts to make lower profile valves with better flow characteristics have resulted in less fibrin deposition on these valves. Thus, the newer generation, low-profile valves may be less predisposed to thrombus formation.

Bioprostheses. Bioprostheses are considerably less thrombogenic mainly because of the natural properties of the materials used and because of their axial flow properties and leaflet pliability. However, bioprosthetic valves are predisposed to thrombus formation in the immediate postoperative period because of platelet deposition and later, in the presence of atrial fibrillation and left atrial enlargement, because of stasis.

In summary, the pathogenesis of arterial thrombosis involves damage to the vessel wall and exposure to a thrombogenic surface leading to both platelet deposition and activation of the coagulation system (Fig. 1). Intracavitary thrombosis occurs primarily in areas where there is stasis which predisposes to activation of the coagulation system over platelet deposition. Finally, mechanical prostheses promote activation of the coagulation system and to a lesser degree platelet activation. Biological prostheses are significantly less thrombogenic.

Risk Stratification of No Antithrombotic Medications: Implications for Antithrombotic Therapy

In the preceding section we described the pathophysiologic basis of thrombus formation in the various clinical syndromes. It is now imperative to consider the various clinical heart disease syndromes according to absolute and relative risk of developing thromboembolic events. We will discuss three general risk categories assuming no antithrombotic medications are used (Table 1): the highest involving more than six episodes per 100 patients per year, a medium risk of 2 to 6 episodes per 100 patients per year, and a lower risk rate of less than 2 events per 100 patients per year. Aggressive antithrombotic management is generally recommended for those with high risk. In contrast, for those at low risk, antithrombotic therapy may not even be recommended. For those in the medium-risk group, an intermediate approach is suggested (13).

Coronary Arteries—High Risk

Acute Coronary Syndromes. In the absence of thrombolytic therapy, the risk of developing an occlusive thrombus in the case of unstable angina, or rethrombosis in the case of lysed thrombus in acute myocardial infarction is high (7).

Coronary Intervention. As mentioned previously, catheter-based interventions lead to platelet activation and activation of the coagulation system with a high risk of acute thrombotic occlusion. Those who have undergone PTCA or stenting have a low, long-term risk of thrombosis (13).

Saphenous Vein Bypass Grafting. As mentioned previously, disease of the saphenous vein graft is the most important factor leading to myocardial ischemia after coronary bypass surgery. Occlusion rates are 8 to 18% per distal anastomosis at 1 month postoperatively and 16 to 26% at 12 months (9). Therefore, within the first year (particularly the first few days) after coronary artery bypass surgery, patients are considered to be at high risk for thrombotic complications. Beyond 1 year patients are considered to be at medium risk for developing thrombotic complications of the saphenous vein graft (13).

Coronary Arteries—Medium Risk

Post-Myocardial Infarction and Chronic Coronary Disease. Survivors of acute myocardial infarction are at medium risk of recurrent infarction or cardiac death. Based on data from the Framingham study, those who have survived a myocardial infarction or unstable angina or who have chronic coronary disease without infarction are estimated to have a risk of 2.5 to 5% of either reinfarction or ischemic death (13).

Coronary Arteries—Low Risk

Primary Prevention of Coronary Events. Individuals who have no risk factors for coronary disease are thought to be at low risk for infarction and ischemic death. However, populations with a high prevalence of coronary disease, i.e., those with many cardiac risk factors, have a higher inherent risk of an acute coronary syndrome and may therefore be in the intermediate risk category.

Cardiac Chambers—High Risk

Mitral Stenosis with Atrial Fibrillation. Atrial fibrillation is a common arrhythmia and is associated with the potentially devastating complication of systemic embolism. The highest risk of systemic thromboembolism in mitral stenosis is in patients with atrial fibrillation. Specifically, the presence of mitral stenosis and atrial fibrillation dramatically increases the risk of thromboembolism and nearly 75% of the patients who have emboli are in chronic atrial fibrillation.

Atrial Fibrillation with a History of Previous Thromboembolism. In patients with atrial fibrillation, the highest risk is in those patients who have experienced a prior thromboembolism. Those with a prior embolism related to mitral stenosis are at high risk for a recurrent thromboembolism. Specifically, the risk of a recurrent event exceeds 10% per year (14).

Cardiac Chambers—Medium Risk

Nonvalvular Atrial Fibrillation. The risk of ischemic stroke in nonvalvular atrial fibrillation is about medium, but is directly related to the presence of coexistent cardiovascular disease. Thus, identification of the subpopulations of patients with higher or lower risk of stroke determines which patients would derive the greatest benefit from anticoagulation.

Two studies, the Stroke Prevention in Atrial Fibrillation Study 1 (SPAF) (15) and the Atrial Fibrillation Study Investigators, provide insight into the risk factors for systemic embolism in chronic or intermittent atrial fibrillation. Left ventricular dysfunction, diabetes mellitus, hypertension, prior stroke or transient ischemic attack, and age over 65 years tend to place patients with atrial fibrillation into the highest risk category. This translates into a risk of 7.5% per year.

Mitral Incompetence in Atrial Fibrillation. This valvular condition is associated with a lower risk of systemic embolism than mitral stenosis (16, 17). As with mitral stenosis, the risk of systemic thromboembolism increases significantly with atrial fibrillation. In a series of more than 800 patients with rheumatic mitral disease studied by Coulshed and colleagues (17), the risk of systemic thromboembolism was 7.7% in those with predominant mitral incompetence with sinus rhythm compared with 22% of those with mitral incompetence and atrial fibrillation (17).

Anterior Myocardial Infarction. Left ventricular thrombus develops in approximately 30% of patients with anterior myocardial infarction and is associated with a medium risk of subsequent arterial embolization (18). Other factors that predict left ventricular thrombus formation are ejection fraction, infarct size, and atrial fibrillation. Thrombus formation almost always forms in the anterior, particularly the anteroapical, area of akinesis or dyskinesis (19, 20). Less than 5% of patients with infarctions in other areas develop a mural thrombus. The risk of embolization from a detected thrombus may be higher when the thrombus is mobile or protuberant (18).

Dilated Cardiomyopathy. In uncontrolled studies, patients with dilated cardiomyopathy have an annual incidence of clinically apparent thromboembolism of 1 to 12 per 100 patients (21, 22). In a study by Fuster et al. (21) patients not receiving anticoagulation had an overall incidence of thromboembolism of 18% or 3.5% per year. This occurred in 14% of patients in sinus rhythm and 33% of patients with atrial fibrillation (21). The risk of thromboembolism is greatest in those with an ejection fraction of less than 35%, established or paroxysmal atrial fibrillation, and a history of thromboembolism or echocardiographic evidence of thrombus (22).

Cardiac Chambers—Low Risk

Lone Atrial Fibrillation. The risk of thromboembolism in lone atrial fibrillation, defined as idiopathic atrial fibrillation in the absence of structural heart disease is low and therefore does not require anticoagulation.

Mitral Valve Disease in Sinus Rhythm. As stated previously, mitral incompetence is associated with a somewhat lower risk of thromboembolism than mitral stenosis. Systemic thromboembolism occurs at a rate of one to four events per 100 patient years, depending on the degree of incompetence and presence or absence of atrial fibrillation (13, 17). Thus, as stated above, in atrial fibrillation the risk is medium whereas in sinus rhythm the risk is low.

The incidence of thromboembolism complicating mitral stenosis in those with normal sinus rhythm is reported as low to medium with a rate of 4.7 to 1.5% per year. This risk increases with age and correlates inversely with cardiac output. It is unrelated to left atrial size, valve area, or functional class. It is important to monitor patients with mitral stenosis for the onset of atrial fibrillation because as many as 30% of the emboli that occur do so within the first month and 65% within the first year after the onset of atrial fibrillation (13).

Aortic Valve Disease. Few long-term studies are available to confirm what is believed to be a low incidence of thromboembolism in aortic valve disease. Isolated aortic valve disease imparts a low risk of thromboembolism, and aortic valve disease associated with atrial fibrillation or calcification of the aortic valve may be associated with embolic events.

Chronic Left Ventricular Aneurysm. Left ventricular aneurysm predisposes to layers of thrombus formation. Intra-aneurysmal thrombi are at high risk for thromboembolism during the first 3 weeks after a myocardial infarction and are at low risk beyond 3 months after myocardial infarction. Specifically, about 5% of patients with intra-aneurysmal thrombi will embolize within 3 weeks after a myocardial infarction (16). Beyond 3 months post-myocardial infarction the incidence of thromboembolism is reported to be 0.3% per year. This lower risk presumably is caused by the presence of an organized thrombus that is somewhat isolated from dynamic forces of the circulation (23).

Prosthetic Hearth Valves—High Risk

Old Mechanical Valves. Without anticoagulation, patients with older mechanical valves are at high risk of thromboembolism. The risk of thromboembolism is cumulative and persistent. Recent observations indicate that even with more modern prosthetic heart valves, there is a medium to high incidence of thromboembolism within the first 30 postoperative days (24).

Mechanical Valves and Previous Thromboembolism. Patients with any type of mechanical prosthesis and a history of thromboembolism appear to be at high risk for recurrent thromboembolic events (24).

Prosthetic Heart Valves—Medium Risk

Newer Mechanical Prostheses. Patients with newer prosthetic valves are at medium risk for thromboembolism. Overall, the risk of thromboembolism has been decreasing in recent years, probably as a result of patient factors, earlier operations, leading to smaller left atrial size, less atrial fibrillation, and better left ventricular function, use of mechanical prosthetic valves with less turbulent flow and better hemodynamics (25).

Bioprosthetic Valves Early Postoperatively or with Atrial Fibrillation. Patients with bioprosthetic valves are at medium risk for thromboembolism within the first 30 days after the operation. In addition, those with any bioprosthetic valves and atrial fibrillation are at chronic medium risk for thromboembolism.

Prosthetic Heart Valves—Low Risk

Bioprosthetic Valves in Normal Sinus Rhythm. Those with bioprosthetic valves and normal sinus rhythm tend to be at low risk for thromboembolic complications beyond the immediate postoperative period.

Antithrombotic Therapy

Because thrombosis is related to platelet and fibrin deposition, the combination of aspirin and anticoagulants may be considered in high-risk conditions (i.e., unstable angina), whereas either aspirin or anticoagulants may be used in medium-risk cases (i.e., chronic coronary disease). In disease conditions related to the cardiac chambers, because thrombotic complications are primarily related to fibrin deposition, the use of anticoagulants at higher doses may be considered in the high-risk conditions (i.e., mitral stenosis with atrial fibrillation), whereas a lower dose may be considered in the medium-risk conditions (i.e., valvular atrial fibrillation). In prosthetic heart valves, because thrombotic complications have a predominance of fibrin over platelet deposition,

TABLE 2

Major Anticoagulant (Heparin) Trials in Unstable Angina

Trial	Follow-up	Drug	Reduction in Death or MI (%)	p
Teleford and Wilson	7 days	Heparin (20,000 U)	80	<0.05
Theroux	6 days	Heparin (24,000 U)	89	<0.001
		Aspirin + heparin	88	0.001
RISC	3 months	Heparin (20,000 U)	5	NS
		Aspirin + heparin	68	<0.0005
ATACS	3 months	Aspirin + heparin	25	0.06

Modified from Freudenberger, Fuster. In: Smith T, editor. Cardiovascular Therapeutics. Boston MA: Saunders, 1996.

the use of anticoagulants at a higher dose or the combination of lower dose anticoagulants and aspirin may be considered in the high-risk conditions (i.e., older mechanical prosthetic valves), whereas a lower dose of anticoagulants alone may be considered in the medium-risk conditions (i.e., recent mechanical prostheses). In this last section we will try to justify such an approach.

Coronary Arteries—High Risk

Unstable Angina (Tables 2 and 3). As stated previously, unstable angina represents a high thrombotic risk because of fibrin deposition and platelet activation. Therefore, therapy is directed toward both systems. The benefit of aspirin and heparin independently has been proven and the use of these agents in combination is strongly suggested.

Heparin Alone. The use of heparin for the treatment of unstable angina was suggested by Teleford and Wilson (26). In a 2 × 2 factorial design study, 400 patients were randomly selected to receive atenolol, heparin, both, or neither. The authors demonstrated an 80% reduction in myocardial infarction in patients with unstable angina treated with heparin for 7 days. In the Montreal Heart Institute Study (27), a double-blind, placebo-controlled 2 × 2 factorial design study, patients were randomly selected to receive heparin alone, aspirin alone, both, or neither. Endpoints of death, myocardial infarction, or severe unstable angina refractory to medical treatment were measured 6 ± 3 days after randomization. Compared with placebo, heparin reduced the total event rate by 57% ($p = 0.001$). The beneficial effects of aspirin and the results of combining aspirin and heparin are discussed below. There was a trend toward less frequent myocardial infarction in the heparin group compared with the aspirin group. An extension of this study was subsequently performed that randomly selected people to receive either aspirin or heparin. The combined results of these studies demonstrated the superiority of heparin over aspirin in reducing the rate of myocardial infarction (28).

Antiplatelet Agents Alone and with Heparin (Tables 2 and 3). Antiplatelet agents have been found to reduce the rate of acute myocardial infarction and short- and long-term death in four large trials of unstable angina (28). In the Veterans Administration Cooperative Study, 1266 men with unstable angina were randomly selected to receive 325 mg/day of aspirin or placebo for 12 weeks. During the treatment period the aspirin group demonstrated a risk reduction of 51%, and the overall benefits of aspirin were maintained during the 1-year follow-up period (29). In the Canadian Multicenter Trial, 55 patients with unstable angina were randomly selected to receive 1300 mg/day of aspirin, 800 mg/day of sulfinpyrazone, the combination of both or placebo. After 18 months the incidence of death and myocardial infarction was reduced in the aspirin groups from 17 to 8.6%; sulfinpyrazone demonstrated no benefit (30).

As mentioned previously, the Montreal Heart Institute Study randomly selected 479 patients with unstable angina to receive 325 mg/twice daily of aspirin, intravenous heparin, the combination of both, or placebo. Compared with placebo, aspirin reduced the rate of myocardial infarction by 72%, and the combination of aspirin and heparin was superior to aspirin alone (27). In the European RISC study group (31), 794 men with unstable angina or non-Q-wave infarction were randomly selected to receive 75 mg/day of aspirin for 3 months, intravenous heparin for 5 days, both or neither. At the end of 3 months the incidence of death or myocardial infarction was significantly reduced by aspirin and, to a greater extent by the combination of aspirin and heparin. The more recent antithrombotic therapy in acute coronary syndromes (ATACS) (32) trial randomly selected 214 patients, who were not prior aspirin users, with either unstable angina or non-Q-wave myocardial infarction, to receive aspirin alone or aspirin plus anticoagulation. The trial was begun

TABLE 3
Major Aspirin Trials in Unstable Angina

Study	Year	No. of Patients	Dose (mg/day)	Duration of Follow-up	Relative Risk Reduction (%)
Lewis et al.	1983	1338	324	3 months	51
Cairns	1985	555	1300	24 months	51
Theroux	1988	479	650	6 days	72
RISC	1990	796	75	5 days	57
				30 days	46

within a mean of 9.5 hours after the onset of chest pain and continued for 12 weeks. Total ischemic events were reduced significantly in the combination group versus aspirin alone (10.5% vs. 27%, $p = 0.004$) at the end of 14 days. At the end of 12 weeks there was a trend toward reduction in total ischemic events in the combination group (13% vs. 25%, $p = 0.06$).

Ticlopidine was examined in patients with unstable angina (33). The administration of ticlopidine in a dose of 250 mg twice daily for 6 months reduced the incidence of death and myocardial infarction by 46%.

Thus, the efficacy of aspirin and heparin in unstable angina is proven by the above-mentioned studies. The benefit of combination therapy, although highly suggested by these studies, has not yet been demonstrated unequivocally relative to monotherapy. However, because of the high risk of thrombotic events and the complementary mechanisms of action in preventing platelet and fibrin deposition, the combination of these two agents for initial treatment of unstable angina has been strongly suggested (34).

Direct Thrombin Inhibitors and IIb/IIIa Platelet Blockers (Table 4). Because patients with unstable angina are at high risk, efforts to develop better antithrombotic medications have been intensified. Given the central role of thrombin in the coagulation process and the role of IIb/IIIa platelet receptors in the activation of platelets, there is great enthusiasm for the development of direct thrombin inhibitors and IIb/IIIa receptor blockers for use in unstable angina. Preliminary results (35) from the OASIS and GUSTO IIB study of the use of the antithrombin hirudin, and the CAPTURE study with the use of IIb/IIIa platelet receptor blockers during the hospitalization phase for unstable angina are encouraging. However, further studies to determine the optimal dosing, route, and length of administration need to be conducted.

Recommendations. In this high-risk group, with the pathophysiologic processes of both platelet activation and fibrin generation, all patients with unstable angina should receive 325 mg of chewable aspirin and intravenous heparin sufficient to raise the partial thromboplastin time to approximately 46 to 70 seconds. Ticlopidine can be used as an alternative to aspirin in those who are allergic or intolerant to aspirin. Evidence of the further beneficial effect of the use of IIb/IIIa platelet receptor blockers or antithrombins instead of heparin is evolving. The use of thrombolytic agents in patients with unstable angina is not recommended.

Acute Myocardial Infarction (Table 5). As stated previously, acute myocardial infarction represents a highly thrombotic state for reocclusion and reinfarction caused by both platelet activation and fibrin generation. Therefore, therapy is directed toward both the platelet and the coagulation system.

Antiplatelet Agents and Clinical Endpoints. (A) Aspirin without thrombolytic therapy for patients with acute myocardial infarction: The landmark Second International Study of Infarct Survival (ISIS-2) (36) which included more than 17,000 patients, randomly selected patients to receive aspirin (160 mg/day for 30 days), intravenous streptokinase, both or neither. In the group taking aspirin alone there was a 23% reduction in mortality compared with the group taking neither streptokinase nor aspirin; the risk reduction for nonfatal reinfarction was 49% and that for nonfatal stroke was 46%. Thus, aspirin is critical in the treatment of acute myocardial infarction.

(B) Treatment with thrombolytic therapy: In ISIS-2, when aspirin was combined with streptokinase, a 42% reduction in vascular mortality was observed. Patients given only aspirin, as stated previously, received a 23% reduction in mortality and whereas those given only streptokinase had a 25% reduction in mortality. Aspirin with streptokinase compared with streptokinase alone reduced the rate of nonfatal infarction (1.8% in the combination group compared with 2.8% in the streptokinase alone group). Thus, the effects of aspirin and thrombolytic agents appear to be additive. No significant difference in major bleeding requiring transfusion was seen. There was a small but significant increase in minor bleeding seen in the aspirin group, however. A recent follow-up study of the ISIS-2 group found that at 4 years there was a persistent reduction in mortality (37).

Anticoagulants and Clinical Endpoints. There have been no clinical trials of aspirin versus heparin during acute myocardial infarction. However, there are studies examining anticoagulants with or without the use of thrombolytic agents, although in the vast majority of these studies patients were receiving aspirin.

(A) Heparin with and without thrombolytic agents: The SCATI (38) group randomly selected 711 patients to receive either heparin or no heparin. Of these patients 433 also received streptokinase. These patients, unlike those in GISSI-2, ISIS-3, and GUSTO (described below), did not receive aspirin. This study found a 44% reduction in mortality when 12,500 units of subcutaneous heparin was given to patients with acute myocardial infarction. This reduction in mortality was significant in both the group receiving streptokinase and in those who did not receive the thrombolytic agent.

(B) Heparin and aspirin with thrombolytic agents: The use of heparin (in those receiving aspirin) as adjunctive therapy with thrombolytic agents was examined in many studies including GISSI-2, ISIS-3, SCATI, and GUSTO. In GISSI-2, patients who received streptokinase and heparin beginning 12 hours after the infusion of the thrombolytic agent had a tendency (but no statistically significant difference) toward a decrease in mortality compared with those who received streptokinase alone. Mortality in those receiving tissue plasminogen activator (t-PA) was the same whether or not heparin was added to the regimen (39). ISIS-3 (40) found that heparin (given subcutaneously, 12,500 units every 12 hours starting 4 hours after the start of thrombolytic therapy) and aspirin, given with t-PA or streptokinase, resulted in a nearly significant decrease in mortality. There was a trend toward a decrease in hospital reinfarction rate in the heparin group. However, this early benefit was lost when the primary endpoint of 35 days was reached. Also, heparin produced a small excess of extracerebral bleeds (1.0% compared with 0.8% ($p < 0.01$) and intracerebral bleeds (0.56% compared with 0.40%; $p < 0.05$).

The GUSTO study provided great insight to the issue of anticoagulation as adjunctive therapy to thrombolysis (41). This study randomly selected 41,021 patients with myocardial infarctions for four treatment groups: 1) streptokinase and subcutaneous heparin, 2) streptokinase and intravenous heparin, 3) accelerated t-PA and intravenous heparin, or 4) the combination of both agents with intravenous heparin. The results showed a significant reduction in mortality in the accelerated t-PA and intravenous heparin group versus the streptokinase groups ($p < 0.001$). The rates of hemorrhagic stroke were 0.49%, 0.54%, 0.72%, and 0.94%, respectively. Because heparin was given in all four groups in conjunction with aspirin, its role given alone or in combination with aspirin was not tested. However, it can be stated that there was no advantage to the use of intravenous over subcutaneous heparin with streptokinase.

Aspirin and/or Anticoagulants with Thrombolytic Therapy and Angiographic Endpoints. Angiographic endpoint studies are inconclusive in nature because they do not provide clinical information. Nevertheless, they can provide some insight into the possible role of certain antithrombotic agents (as adjuncts to thrombolytic therapy) not tested in terms of clinical endpoints, such as heparin versus aspirin, versus the combination of both. Several smaller studies of angiographic patency examined the use of heparin in conjunction with t-PA. The Heparin Aspirin Reperfusion Trial (HART) compared heparin with 80 mg of aspirin per day after administration of t-PA (42). Coronary artery patency at 18 hours was 82% in the heparin group compared with 52% in the aspirin-alone group ($p < 0.002$). Bleich and colleagues (43) found that, in patients not receiving aspirin (80 mg/day) during acute myocardial infarction after administration of t-PA, patency (as assessed by angiography) of the infarct-related artery at 2 to 3 days was improved by heparin therapy (71% vs. 44%, $p = 0.04$). In the European Cooperative Study Group-6, 687 patients received aspirin and t-PA. Half of this group was randomly selected to receive heparin, whereas the other half received no heparin. Patency at 81 hours was 80% in the heparin group and 75% in the no-heparin group ($p < 0.01$) (44). The Australian National Heart Study determined that there was no difference in patency at 1 week between patients receiving continuous heparin infusion after an initial bolus or those receiving a combination of 300 mg of aspirin and 300 mg of dipyridamole after the initial heparin bolus after t-PA infusion (45).

In a recent study, the Duke University Clinical Cardiology Studies-1 found no difference in coronary artery patency in those receiving anisoylated plasminogen streptokinase activator complex (APSAC) and heparin versus APSAC alone (46). All patients were given aspirin.

TABLE 4

New Antithrombotic Agents/In Acute Coronary Syndromes: Clinical Trials

Agents	Unstable Angina	Myocardial Infarction	PTCA Occlusion
Hirudin	GUSTO IIA^{a} *X* GUSTO IIB^{g}*X* OASIS	GUSTO-IIA^{a}*X* TIMI-9A^{b}*X* HIT III X GUSTO-IIB^{g}*X* TIMI 9B X OASIS	HELVETICA X
Hirulog	TIMI-8 X	HERO* (SK)^{f}	PTCA X
IIb/IIIa blockers	CAPTURE *X* PURSUIT ? PRISM ? PARAGON ?	IMPACT MI ? PARADIGM ?	IMPACT II X PTCA^{d}*X* RESTORE^{f}*X* EPIC^{e}* (& restenosis) EPILOG^{f}* ERASER (STENT) ?

X = results of trial questionable; ? = trial in progress; * = trial showed definitive benefit; asa = aspirin; a/c = anticoagulation.

[a] The GUSTO Investigators. Randomized trial of intravenous heparin versus recombinant hirudin for acute coronary syndromes (GUSTO-IIa). Circulation 1994; 90:1631–1637.

[b] Antman E. Hirudin in acute myocardial infarction (TIMI-9A). Circulation 1994; 90:1624–1630.

[c] Neuhaus K, von-Essen R, Tebbe U, et al. Safety observations from the pilot phase of the randomized r-hirudin in thrombolysis study. Circulation 1994; 90:1638–1642.

[d] Tcheng J, Harrington R, Kottke-Marchant K, et al. Multicenter, randomized, double blind, placebo controlled trial of platelet integrin glycoprotein IIb/IIIa blocker, integrelin in elective coronary intervention. Circulation 1995; 91:2151–2155.

[e] Epic Investigators. Use of monoclonal antibodies directed against the platelet glycoprotein IIb/IIIa receptor in high risk coronary angioplasty. N Engl J Med 1994; 330:956–961.

[f] Oral presentation at the American College of Cardiology national meeting 1996.

[g] Gusto IIB Investigators. A comparison of recombinant hirudin versus heparin in the treatment of acute coronary syndromes. N Engl J Med 1996; in press.

Direct Thrombin Inhibitors and IIb/IIIa Platelet Receptor Blockers (***Table 4***). Because patients with acute myocardial infarction are at high risk, efforts to develop better antithrombotic agents have been intensified. Because thrombin has a central role in the coagulation process and IIb/IIIa platelet receptors have a central role in the activation of platelets, as already indicated in unstable angina, there is great enthusiasm for the development of direct thrombin inhibitors and IIb/IIIa receptor blockers to be used in acute myocardial infarction (35). Preliminary results from the OASIS study using the antithrombin hirudin and the HERO study using the antithrombin Hirulog are very encouraging. However, further studies to determine the optimal dosing, route, and length of administration and the optimal protocol need to be conducted.

Recommendations. In summary, the efficacy of aspirin and the efficacy of heparin in acute myocardial infarction has been proven. However the benefit of combination therapy with thrombolytic therapy versus monotherapy using either aspirin or heparin is suggested if t-PA is used but has not yet been demonstrated unequivocally; at the present time, no study supports the use of either intravenous or subcutaneous heparin as adjunctive therapy to streptokinase when an adequate dose (162–325 mg/day) of aspirin is used. The same probably applies to the use of heparin in conjunction with APSAC. Based on the data obtained with unstable angina, in myocardial infarction without thrombolytic therapy, the in-hospital use of the combination of aspirin and heparin is justified; however, this approach has not yet been tested. With the use of thrombolytic therapy, heparin should be used in conjunction with aspirin and in patients who would otherwise benefit from anticoagulation for preventing systemic embolism, i.e., those with a large anterior myocardial infarction, cardiac failure, or atrial fibrillation. The further beneficial effect of the use of IIb/IIIa platelet receptor blockers or new antithrombins has yet to be elucidated.

Coronary Intervention. Based on several major trials, pretreatment with aspirin combined with adequate heparinization throughout the procedure is strongly recommended (47). Ticlopidine can be used as an alternative to aspirin in those patients who are allergic or intolerant to aspirin. Trials of the use of different antiplatelet drugs and anticoagulant agents aimed at reducing the restenosis rate have been disappointing. However, use of the platelet glycoprotein receptor IIb/IIIa antagonist has recently received much attention. The EPIC study showed reduced incidence of acute and delayed complications, including restenosis after PTCA, in the group of patients receiving the 7E3 antibody directed against the glycoprotein IIb/IIIa receptor complex (48, 49). This approach is currently recommended in all patients considered high risk for ischemic complications after coronary intervention.

In patients undergoing stent implantation, use of aspirin and heparin during the procedure are presently recommended. Subsequent treatment regimens (less aggressive) using aspirin alone or in combination with ticlopidine has been suggested, particularly in conjunction with the use of intravascular ultrasound guidance to guide deployment of the stent (50).

Saphenous Vein Bypass Grafting. Reviews of all reported trials to date have demonstrated the importance of initiating platelet inhibitor therapy in the perioperative period, preferably before 48 hours after surgery. When therapy was not started 48 hours after surgery, no reduction in the vein graft occlusion rate was observed (9, 51).

In the Veterans Administration Cooperative Study, patients receiving saphenous vein grafts were randomly assigned into five groups, taking 325 mg aspirin daily, 325 mg three times daily, 325 mg aspirin plus 75 mg dipyridamole three times daily, 267 mg sulfinpyrazone three times daily, or placebo. Early graft patency at a median of 9 days was significantly higher in the aspirin-treated group (92%) than in those receiving placebo (85%). At 1 year benefit was seen in patients at high risk of graft closure (those with grafts placed to vessels < 1.5 mm in diameter) taking aspirin. One daily dose of aspirin was as effective as three daily doses. Dipyridamole conferred no benefit over aspirin alone. In another Veterans Administration study, aspirin used preoperatively was associated with increased bleeding complications and provided no additional benefit over aspirin started 6 hours after surgery (52). In a study examining the use of ticlopidine 250 mg twice daily started on the second postoperative day, there was a 40% reduction in vein graft occlusion, as assessed angiographically at 10, 180, and 360 days after surgery (53).

Currently, aspirin should be started immediately after surgery and continued for at least 1 year and probably indefinitely. Ticlopidine can be used as an alternative to aspirin. However, because the occlusion rate remains high, combination therapy should be tested in the future.

Post-Myocardial Infarction-Chronic Coronary Artery Disease (Table 5). As stated previously, survivors of acute myocardial infarction are at medium risk of recurrent infarction or cardiac death. Because morbidity and mortality after a myocardial infarction may be related to arrhythmias, left ventricular dysfunction, and recurrent myocardial infarction, proving that antithrombotic therapy is beneficial in these patients has been difficult.

Antiplatelet Agents. Multiple trials examining the use of aspirin for the secondary prevention of myocardial infarction have been conducted, but no single study has provided definitive results. The results of a meta-analysis that included more than 18,000 patients revealed that platelet inhibitor therapy reduced cardiovascular mortality by 13%, nonfatal reinfarction by 31%, and nonfatal stroke by 42% (53). Aspirin alone was at least as effective as the combination of aspirin and dipyridamole and more effective than sulfinpyrazone. Available data do not justify the additional cost and frequency of administration of drugs other than aspirin in this group of patients (54). Medium-dose aspirin (75–325 mg) was as efficacious as higher-dose aspirin (55). No studies on the use of ticlopidine for secondary prevention are available.

Anticoagulants. Numerous studies have assessed the usefulness of anticoagulants in the secondary prevention of cardiovascular disease after myocardial infarction. The Warfarin and Reinfarction Study (WARIS) is the largest study to date of anticoagulants in the secondary prevention of myocardial infarction (56). In this placebo-controlled, double-blind trial, 1214 patients aged 75 years or less were randomly selected a mean 27 days after initial acute myocardial infarction (AMI) to receive warfarin (target international normalized ratio [INR] 2.8–4.8) or placebo. Only 1% of patients received thrombolytic therapy for the initial AMI. Patients were advised not to take aspirin during the trial. At a mean follow-up of 37 months, use of warfarin had resulted in significant reduction of mortality, total reinfarctions, and total strokes. Although four fatal intracranial hemorrhages occurred in the warfarin group compared with none in the placebo group, 10 nonhemorrhagic fatal strokes occurred in the placebo group compared with none in the warfarin group. Efficacy analysis showed even greater benefit to warfarin.

Two clinical trials compared warfarin therapy with antiplatelet therapy in secondary prevention of myocardial infarction. In the German Austrian Myocardial Infarction Study (GAMIS), 946 patients were randomly selected 38 to 42 days post-AMI to receive open-label phenprocumin (target INR 2.5–5.0), aspirin 1.5 g/day, or placebo (57). There was no difference observed between

TABLE 5

Aspirin (ASA) and Anticoagulant Agents (A/C) Efficacy in CAD Patients: Secondary Prevention

	MI <1 Month				MI >1 Month–3 yr		Chronic Stable CAD	
	No lysis		Lysis					
Events	ASA[a]	A/C[b]	ASA[a]	A/C[c]	ASA[d]	A/C[e,f]	ASA[g–i]	A/C[j]
Nonfatal CVA	46	51	50	–	42	63	25	–
Nonfatal MI	49	22	53	–	31	38	39	56
Death	23	16	42	–	13	22	26	26

[a] ISIS II Collaborative Group. Randomized trial of intravenous streptokinase, oral aspirin, both or neither among 17,187 cases of suspected acute myocardial infarction. Lancet 1988; 2:349–360.

[b] MacMahon S, Collins R, Knight C, et al. Reduction in major morbidity and mortality by heparin in acute myocardial infarction. Circulation 1988; 78(II):98.

[c] Husia J, Hamilton W, Kleiman N, et al. A comparison between heparin and low dose aspirin as adjunctive therapy for acute myocardial infarction (HART). N. Engl J Med 1990; 323:1433–1437.

[d] Antiplatelet Trialist's Collaboration. Collaborative overview of randomized trials of antiplatelet therapy. Br Med J 1994; 308:81–106.

[e] Smith P, Arnesen H, Holme I. The effect of warfarin on mortality and reinfarction after myocardial infarction. N Engl J Med 1990; 1323:147–152.

[f] The ASPECT Research Group, The effect of long term anticoagulant treatment on mortality and cardiovascular morbidity after myocardial infarction. Lancet 1994; 343:499–503.

[g] SAPAT Group. Double blind trial of aspirin in primary prevention of myocardial infarction. Lancet 1992; 340:1421.

[h] Chesebro J, Webster M, Smith H, et al. Antiplatelet therapy in coronary disease progression. Circulation 1989; 80(II):266.

[i] Steering Committee of the Physicians Health Study. Final report on the aspirin component of the ongoing physician's health study. Ann Intern Med 1991; 114:835.

[j] Sixty Reinfarction Study Group. A double blind trial to assess long-term oral anticoagulant therapy in elderly patients. Lancet 1980; 2:989–984.

Modified from Freudenberger. Fuster. In: Cardiovascular therapeutics. Smith T, editor. Boston MA: Saunders, 1996.

groups in mortality or reinfarction. The French Enquete de Prevention Secondaire de l'Infarctus de Myocarde (EPSIM) study (58) revealed no difference in death or reinfarction in patients receiving either oral anticoagulants or aspirin. However, there were 54% more patients with gastrointestinal events with aspirin and four times more severe hemorrhagic events with warfarin. In the Aspirin versus Coumadin in the Prevention of Reocclusion and Recurrent Ischemia After Successful Thrombolysis (APRICOT) trial, 300 patients were randomly selected to receive either 325 mg of aspirin per day or to receive heparin followed by warfarin (target INR 2.8–4.0) after an initial angiogram <48 hours post-AMI revealed a patent infarct-related artery (59). At 3 months there was no significant difference in reocclusion rates among the warfarin, aspirin, and placebo arms. Aspirin significantly reduced reinfarction compared with placebo but not with warfarin. Mortality did not differ among the groups.

The combination of anticoagulants and antiplatelet agents in the secondary prevention of acute myocardial infarction has been examined recently. Results of the recent CARS (Coronary Artery Reinfarction Study) (60), which planned to study 9000 patients randomly selected for three treatment regimens: 1) 160 mg of aspirin per day; 2) 80 mg of aspirin per day *plus* 3 mg of warfarin per day; 3) a combination pill of 80 mg of aspirin plus 1 mg of warfarin per day, were recently reported. This trial was stopped after an interim analysis that revealed that the combination of fixed low-dose coumadin plus low-dose aspirin is no more effective than aspirin in the prevention of reinfarction, cardiovascular death, and stroke.

The CHAMP study will randomly select 4000 patients to receive either 160 mg of aspirin per day or 80 mg of aspirin plus coumadin to achieve an INR of 1.5–2.5. Results of these studies may definitively settle some of the unanswered questions regarding the use of anticoagulants and antiplatelets for the secondary prevention of AMI.

Primary Prevention of Acute Myocardial Infarction. Two studies examining the use of aspirin in primary prevention of myocardial infarction were conducted. In the United States, for the Physician's Health Study more than 22,000 male physicians aged 40 to 84 years were randomly selected to receive 325 mg of aspirin every other day or placebo for 5 years. There was a 44% reduction in acute myocardial infarction in the aspirin-treated group. This effect was limited to those older than 50 years. The incidence of cardiovascular death was not different (61).

In the British primary prevention trial of more than 5000 male physicians aged 50 to 78 years, two thirds were randomly selected to receive 500 mg of aspirin per day and one third was advised to use no aspirin. After 6 years there was no difference in the rate of myocardial infarction (62). These two studies evaluated the use of antiplatelet agents in a group that had lower risk than the general population; the use of these agents may in fact be beneficial in the high-risk population. Therefore, many physicians recommended the use of aspirin in the high-risk patient for the primary prevention of the acute coronary syndromes and no therapy in the low-risk population.

Cardiac Chambers

Antithrombotic therapy in the cardiac chambers is directed primarily against the deposition of fibrin and consists of antithrombotic therapy. The high-risk situations of atrial fibrillation and a history of thromboembolism and atrial fibrillation and mitral stenosis require a high degree of anticoagulation. Atrial fibrillation and left ventricular dysfunction after an anterior myocardial infarction are medium-risk situations and require a lesser degree of anticoagulation.

Atrial Fibrillation. Because of the pathophysiologic basis of thrombosis in the cardiac chambers, predominantly a fibrin-mediated event, the use of anticoagulants have been studied extensively. Five randomized, clinical trials using warfarin for the prevention of stroke in atrial fibrillation were reported recently (63). These studies demonstrated that an INR of 1.8–4.2 translates to a mean reduction in the rate of ischemic stroke of nearly 70%. This reduction is even more apparent in those patients with high risk of thromboembolism. The incremental risk of bleeding in those followed closely in these studies was <1% per year (63). Safety in the older age group (age >75 years) has not been assessed adequately. One study analyzing safety with patients aged 75 years demonstrated a withdrawal rate of 38.5% per year. A recent study comparing patients older than 75 years of age who had atrial fibrillation to younger patients who had atrial fibrillation found that the very elderly had a greater risk of stroke and hemorrhage on full-dose anticoagulation with an INR of 2.7 (63). It is conceivable that a lower intensity of anticoagulation in this age group may translate into less stroke and lower rates of bleeding complications (64). This approach is currently being evaluated in several clinical trials.

Pooled data examining the efficacy of aspirin in atrial fibrillation patients appears to result in risk reduction of approximately 25%. However, aspirin was significantly less effective than anticoagulation in these studies (15). Thus, we currently recommend that low-risk patients younger than 65 years of age should receive 325 mg of aspirin. Those who are at higher risk who can safely receive anticoagulation should receive warfarin with a target INR 2.0–3.0. Those older than 75 years of age should receive coumadin and their INR levels watched closely because of the apparent greater risk of bleeding complications (65).

As mentioned previously, those who have experienced a thromboembolic event from atrial fibrillation are at a risk exceeding 10% per year of developing a recurrent event. A recent large randomized trial showed that anticoagulation (INR 2.5–4.0) was highly effective and relatively safe in the secondary prevention of stroke in atrial fibrillation (66). Because of the high rate of recurrent stroke, these patients should begin receiving anticoagulation within 1 to 2 weeks of stroke (depending on the size of the stroke and the risk of hemorrhage) to target INR 2.5–3.5. Those who are unable to receive anticoagulants should receive aspirin, which confers some benefit in secondary prevention (66).

Valvular Heart Disease. *Mitral Stenosis.* No prospective randomized studies have evaluated the effectiveness of anticoagulation in mitral stenosis. However, several descriptive studies report that warfarin is effective in preventing recurrent emboli. Thus, patients with mitral stenosis and atrial fibrillation and those patients with valvular heart disease and a prior thromboembolic event are at a higher risk and should receive anticoagulations with an INR of about 2.5–3.5.

Mitral Incompetence. Patients with mitral incompetence or mixed lesions and atrial fibrillation should receive long-term anticoagulation therapy to maintain the INR at 2 to 3.

Aortic Valve Disease. As mentioned previously, aortic valve disease probably represents a low risk for thromboembolism and does not warrant anticoagulation.

Left Ventricular Thrombus. The use of heparin in acute myocardial infarction may prevent the occurrence of left ventricular thrombus formation. The effect of subcutaneous heparin on the incidence of intracardiac thrombosis was evaluated in two clinical trials in which patients receiving 12,500 units every 12 hours were compared with either an untreated group or patients receiving 5000 units subcutaneously every 12 hours. In these two studies the incidence of mural thrombosis detected by 2D-echo was 72% and 58%

lower, respectively, in patients taking a moderate dose (12,000 units) of heparin (38, 67).

Heparin should be used in all anterior myocardial infarctions, with long-term anticoagulation (3 months) in patients with large anterior myocardial infarctions and with patients who have visible thrombi at echocardiography. Use of long-term anticoagulation in patients with postinfarction cardiomyopathy is controversial. Use of anticoagulation in patients with ventricular aneurysms and mural thrombi is generally not indicated because of the low risk of embolization.

Prosthetic Heart Valves

Mechanical Valves. As stated previously, thromboembolic complications of prosthetic heart valves are primarily related to fibrin deposition. The high-risk categories are old prosthetic valves and prosthetic valves with prior thromboembolism. These categories require a high degree of anticoagulation. Recent mechanical prosthesis and bioprosthesis with atrial fibrillation are situations that represent medium risk and require an intermediate degree of anticoagulation.

The optimal anticoagulation regimen has not been fully defined. Recent retrospective studies have questioned the generally recommended regimen of an INR of 3–4.5 for aortic and mitral valve prostheses (24, 68). Thus, for mechanical prostheses in the aortic and mitral position an INR of 2.5–3.5 may be adequate. An INR of 2–3 may suffice with bioprosthetic valves in the presence of atrial fibrillation and no anticoagulation at all with bioprosthetic valves and normal sinus rhythm. A large-scale prospective randomized trial (GELIA) is underway to determine the optimal level of anticoagulation after valve replacement with the St. Jude prosthesis.

No randomized placebo-controlled study has been conducted using platelet inhibitors alone for prevention of thromboembolism with mechanical valve prostheses. One study investigated the combination of dipyridamole and aspirin or pentoxifylline and aspirin and found a significantly higher rate of thromboembolism in these groups when compared with those receiving warfarin therapy in this open prospective randomized trial of patients with a mitral or aortic Starr-Edwards valve (69). Other small studies examining either aspirin alone or in combination with other antiplatelet agents have failed to demonstrate adequate protection from thromboembolism.

Although consistent anticoagulation for patients with mechanical heart valves provide a decreased incidence of thromboembolism, there is still a relatively high risk of residual thromboembolism. Several trials have examined if the addition of dipyridamole to warfarin prevents more episodes of thromboembolism than warfarin alone. In three trials the combination of warfarin and dipyridamole resulted in a 70 to 90% reduction in embolic events. In two other trials the addition of dipyridamole to warfarin resulted in a 40 to 50% reduction in embolic events. Because of the low event rates in these trials these results were not statistically significant (25). The addition of aspirin in doses of 500 to 1000 per day to warfarin lead to a significant decrease in embolic events but also led to a significant increase in hemorrhagic events (25).

Thus, subcutaneous heparin or intravenous heparin should be considered once the chest tubes are removed and continued until oral anticoagulation has reached the desired intensity. An INR of 2.5–3.5 is currently recommended. Because the consistency of anticoagulation is critical, we recommend monitoring every 3 to 4 weeks. High-risk patients, i.e., those with a mechanical valve prosthesis implanted before 1975 and those with prior thromboembolic complications should have a target INR of 3–4.5 or medium intensity anticoagulation (INR 2.5–3.5) and the use of a platelet inhibitor, either 100 mg of aspirin per day or 100 mg of dipyridamole four times daily (16). Platelet inhibitor therapy alone is not advisable in patients with mechanical heart valve prostheses. In patients who develop an absolute contraindication to anticoagulation after implantation of a mechanical valve prosthesis, the combination of 100 mg of dipyridamole four times per day and 325 mg of aspirin daily or 250 mg of ticlopidine twice daily may be tried empirically.

Bioprosthetic Valves. Patients with biological heart valve prostheses should receive low-intensity, oral anticoagulation to an INR 2.0–3.0 beginning as soon as possible after surgery and continuing 3 months postoperatively. Patients with bioprostheses who are at high risk for thromboembolism, including those with atrial fibrillation and previous thromboembolism, should receive oral anticoagulation indefinitely (16).

Conclusion

Understanding the pathogenesis and risk of thrombus formation, a rational approach to the use of antiplatelet agents, antithrombotic agents, and potentially direct thrombin inhibitors can be formulated. Those at high risk of thrombus formation should generally receive a high degree of anticoagulation and, depending on the pathophysiology of the thrombus, may also benefit from the concomitant use of antiplatelet agents. Those with a medium risk of thrombus formation may benefit with less anticoagulation or the use of antiplatelet agents alone. Patients at low risk of thrombus formation should not receive anticoagulation.

References

1. McLean J. The thromboplastic action of cephalin. Am J Physiol 1916; 41:250–257.
2. Fuster V, Dyken M, Vokonas P, et al. AHA medical/scientific report. Aspirin as a therapeutic agent in cardiovascular disease. Circulation 1993; 87:659–675.
3. Stahmann M, Huebner C, Link K. Studies on hemorrhagic sweet clover disease: identification and synthesis of the hemorrhagic agent. J Biol Chem 1941; 138: 513–517.
4. Fletcher A, Sherry S, Alkajaersig N, et al. The maintenance of a sustained thrombolytic state in man. II: Clinical observations on patients with myocardial infarction and thromboembolic disorders. J Clin Invest 1959; 38:1111–1119.
5. Veterans Administration Cooperative Study. Anticoagulants in acute myocardial infarction: results of a cooperative clinical trial. JAMA 1973; 225:724–729.
6. Drapkin A, Merskey C. Anticoagulants after acute myocardial infarction. JAMA 1972; 222:541–548.
7. Fuster V, Badimon L, Badimon J, Chesebro J. The pathogenisis of coronary artery disease and the acute coronary syndromes. N Engl J Med 1992; 326:242–250,310–318.
8. Dangas G, Fuster V. Management of restenosis after coronary intervention. Am Heart J 1996; 132 (Part 1):428–436.
9. Fuster V, Chesebro J. Role of platelets and platelet inhibitors in aortocoronary artery vein-graft disease. Circulation 1986; 73:227–232.
10. Ambrose J, Tannenbaum M, Alexopoulos D, et al. Angiographic progression of coronary artery disease and development of myocardial infarction. J Am Coll Cardiol 1988; 12:56–62.
11. Virchow R. Gesammelte Abhandlungen zur Wissenschaftlichen Medicine. Frankfurt: Meidinger Sohn & Co. 1856: 219–732.
12. Dewanjee M, Fuster V, Rao S, et al. Noninvasive radioisotopic method for detection of platelet deposition in mitral valve prostheses and quantification of visceral microembolism in dogs. Mayo Clin Proc 1983; 58:307–314.
13. Stein B, Fuster V, Halperin J, et al. Antithrombotic therapy in cardiac disease: an emerging approach based on pathogenesis and risk. Circulation 1989; 80:1501–1513.
14. Stroke Prevention in Atrial Fibrillation Investigators. Predictors of stroke in atrial fibrillation II: An echocardiographic features of patients at risk. Ann Intern Med 1992; 116:6–12.
15. Stroke Prevention in Atrial Fibrillation Investigators. The stroke prevention in atrial fibrillation study: final results. Circulation 1991; 84:257.
16. Israel D, Fuster V, Ip J, et al. Intracardiac thrombosis and systemic embolization. In: Colman R, Hirsh J, Marder V, et al., editors. Hemostasis and thrombosis: basic principles and clinical practice. 3rd ed. Philadelphia: J. B. Lipincott Company, 1994: 1452–1468.
17. Coulshed N, Epstein E, McKendrick C, et al. Systemic embolism in mitral valve disease. Br Heart J 1970; 32: 26.
18. Kontny F, Dale J, Hegren L, et al. Left ventricular thrombosis and arterial embolism after thrombolysis in acute anterior myocardial infarction: predictors and effects of adjunctive antithrombotic therapy. Eur Heart J 1993; 14:1489–1492.
19. Visser C, Kan G, David K, et al. Two-dimensional echocardiography in the diagnosis of left ventricular thrombus. Chest 1983; 83:228–232.
20. Penny W, Chesebro J, Heras M, Fuster V. Antithrombotic therapy for patients with cardiovascular disease. Curr Prob Cardiol 1988; 13:464–469.
21. Fuster V, Gersh B, Giuliani E, et al. The natural history of idiopathic dilated cardiomyopathy. Am J Cardiol 1981; 47:525–531.
22. Falk R, Foster R, Coats M. Ventricular thrombi and thromboembolism in dilated cardiomyopathy: a prospective follow up study. Am Heart J 1992; 123:136–142.
23. Lapeyere I, Steele P, Kazimer F. Systemic embolism in chronic ventricular aneurysm: incidence and the role for anticoagulation. J Am Coll Cardiol 1985; 6:534.
24. Heras M, Chesebro J, Fuster V, et al. High risk of thromboembolism early after bioprosthetic cardiac valve replacement. J Am Coll Cardiol 1995; 25:1111–1119.
25. Israel D, Sharma S, Fuster V. Antithrombotic therapy in prosthetic heart valve replacement. Am Heart J 1994; 127:400–411.
26. Teleford A, Wilson C. Trial of heparin versus atenolol in prevention of myocardial infarction in intermediate coronary syndromes. Lancet 1981; 1:1225.
27. Theroux P, Ouimet H, McCanu T. Aspirin, heparin or both to treat acute unstable angina. N Engl J Med 1988; 319:1105.
28. Hirsh J, Fuster V. AHA Medical/Scientific Statement Guide to Anticoagulant Therapy. Part 1: Heparin. Circulation 1994; 89:1449–1468.
29. Lewis H, Davis J, Archibald D, et al. Protective effects of aspirin against acute myocardial infarction and death in men with unstable angina: results of a Veterans Administration Cooperative Study. N Engl J Med 1983; 309:396–403.
30. Cairns J, Gent M, Singer J, et al. Aspirin sulfinpyrazone or both in unstable angina. N Engl J Med 1985; 313: 1369–1375.
31. RISC Investigators: Risk of myocardial infarction and death during treatment with low-dose aspirin and intravenous heparin in men with unstable coronary artery disease. Lancet 1990; 336:827–830.
32. Cohen M, Adams P, Parry G, et al. Combination antithrombotic therapy in unstable rest angina and non-Q-wave infarction in nonprior aspirin users. Primary end points analysis from the ATACS trial. Circulation 1994; 89:81–88.

33. Balsano F, Rizzon P, Violoi F, et al. Antiplatelet treatment with ticlopidine in unstable angina: A controlled multicenter clinical trial. The Studio della Ticlopidina nell'Angina Instabile Group. Circulation 1990; 82:17–26.
34. Braunwald E, Jones R, Mark D, et al. Diagnosing and managing unstable angina. Circulation 1994; 90:613–622.
35. Coller BS. Platelets and thrombolytic therapy. N Engl J Med 1990; 322:33–42.
36. ISIS-2 Collaborative Group. Randomized trial of intravenous streptokinase, oral aspirin, both or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. Lancet 1988; 2:349–360.
37. Baigent C, Collins R, for the ISIS Collaborative Group. ISIS-2: 4 year mortality follow-up of 17187 patients after fibrinolytic and antiplatelet therapy in suspected acute myocardial infarction. Circulation 1993; 88(4 Part 2):I-291. Abstract 1558.
38. The SCATI Group. Randomized controlled trial of subcutaneous calcium heparin in acute myocardial infarction. Lancet 1989; 2:182–186.
39. GISSI-2. A factorial randomized trial of alteplase versus streptokinase and heparin versus no heparin among 12,490 patients with acute myocardial infarction. Lancet 1990; 336:65–71.
40. ISIS-3 Collaborative Group. ISIS-3: a randomized comparison of streptokinase vs t-PA vs anistreplase and of aspirin plus heparin vs aspirin alone among 41,229 cases of suspected acute myocardial infarction. Lancet 1992; 339:753–770.
41. The GUSTO Investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. N Engl J Med 1993; 329:678–682.
42. Hsia J, Hamilton W, Kleiman N, et al. A comparison between heparin and low dose aspirin as adjunctive therapy with tissue plasminogen activator for acute myocardial infarction: Heparin-Aspirin Reperfusion Trial (HART) Investigators. N Engl J Med 1990; 323:1433–1437.
43. Bleich S, Nichols T, Schumacher R, et al. The effect of heparin on coronary arterial patency after thrombolysis with tissue plasminogen activator in acute myocardial infarction. Am J Cardiol 1990; 66:1412–1417.
44. de Bono D, Simoons M, Tijssen J, Arnold A, European Cooperative Study Group. Effect of early intravenous heparin on coronary patency, infarct size, and bleeding complications after ateplasse thrombolysis: results of double blind European Cooperative Study Group trial. Br Heart J 1992; 67:122–128.
45. Australian Coronary Thrombolysis Group. A randomized comparison of oral aspirin/dipyridamole versus intravenous heparin after rt-PA for acute myocardial infarction. Circulation 1989; 80s:114.
46. O'Conner C, Meese R, Camey R. A randomized trial of intravenous heparin in conjunction with APSAC in acute myocardial infarction: the Duke University Clinical Cardiological Study (DUCCS). J Am Coll Cardiol 1994; 23:11–18.
47. Popma J, Coller B, Ohman E, et al. Antithrombotic therapy in patients undergoing coronary angioplasty. Chest. In press.
48. EPIC Investigators. Use of monoclonal antibody directed against the platelet glycoprotein IIb/IIIa receptor in high risk coronary angioplasty. N Engl J Med 1994; 330:956–961.
49. Topol E, Califf R, Weisman H. Randomised trial of coronary intervention with antibody against platelet glycoprotein IIb/IIIa integrin for reduction of clinical restenosis: results at six months. Lancet 1994; 343:881–886.
50. Colombo A, Hall P, Nakamura S, et al. Intracoronary stenting without anticoagulation accomplished with intravascular ultrasound guidance. Circulation 1995; 91: 1676–1688.
51. Chesebro J, Goldman S. Coronary artery bypass surgery. In: Fuster V, Verstreate M, editors. Thrombosis in cardiovascular disorders. Philadelphia: WB Saunders, 1992: 375–388.
52. Goldman S, Copeland J, Moritz T, et al. The Department of Veteran Affairs Cooperative Study Group. Starting aspirin therapy after operation: Effects on early graft patency. Circulation 1991; 84:520–526.
53. Limet R, David J, Magotteaux P. Prevention of aortocoronary bypass graft occlusion: Beneficial effect of ticlopidine on early and late patency of venous coronary bypass grafts. A double blind study. J Thor Cardiovasc Surg 1987; 94:773–783.
54. Antiplatelet Trialists Collaborative Secondary prevention of vascular disease by prolonged antiplatelet treatment. Br Med J 1988; 296:320.
55. Antiplatelet Trialists Collaboration. Collaborative overview of randomized trials of antiplatelet therapy: Parts I, II, III. Br Med J 1994; 308:159–235.
56. Smith P, Arnesen H, Holme I. The effect of warfarin on mortality and reinfarction after myocardial infarction. N Engl J Med 1990; 323:147.
57. Breddin D, Loew D, Lechner K, et al. The German-Austrian Aspirin trial: a comparison of aspirin, placebo and phenprocoumon in secondary prevention of myocardial infarction. Circulation 1980; 62:63.
58. The EPSIM Research Group. A controlled comparison of aspirin and oral anticoagulants in prevention of death after myocardial infarction. N Engl J Med 1982; 307: 701.
59. Meijer A, Verheug F, Werter C. Aspirin versus coumadin in the prevention of reocclusion and recurrent ischemia ater successful thrombolysis: a prospective placebo-controlled angiographic study: results of the APRICOT study. Circulation 1993; 87:1524.
60. Fuster V., for the CARS Investigators. Presented to the American College of Cardiology. March 1996.
61. Steering Committee of the Physician's Health Study Research Group. Final report on the aspirin component of the ongoing Physician's Health Study. N Engl J Med 1989; 321:129–135.
62. Peto R, Gary R, Collins R, et al. A randomized trial of the effects of the effects of prophylactic daily aspirin among male British doctors. Br Med J 1988; 313–316.
63. Atrial Fibrillation Investigators. Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation. Analysis of pooled data from five randomized trials. Ann Intern Med 1994; 154:1449.
64. Veterans' Affairs Stroke Prevention in Nonrheumatic Atrial Fibrillation Investigators. Warfarin in the prevention of stroke associated with nonrheumatic atrial fibrillation. N Engl J Med 1992; 327:1406.

65. Prytowsky E, Benson W, Fuster V, et al. Management of patients with atrial fibrillation: a statement for healthcare professionals. Circulation 1996; 93:1262–1277.
66. EAFT Study Group. European atrial fibrillation trial: secondary prevention of vascular events in patients with non rheumatic atrial fibrillation and recent transient ischemic attack or minor stroke. Lancet 1993; 342:1255.
67. Turpie A, Robinson J, Doyle D. A comparison of high dose with low dose subcutaneous heparin to prevent left ventricular mural thrombosis in patients with acute transmural anterior myocardial infarction. N Engl J Med 1989; 320:352–357.
68. Piper C, Schulte H, Horstkotte D. Optimization of oral anticoagulation for patients with mechanical heart valve prosthesis. J Heart Valve Dis 1995; 4:127–137.
69. Butchart E. Rationalizing antithrombotic management for patients with mechanical heart valves. J Heart Valve Dis 1995; 4:106–113.

Congress and the International Medical Graduate:

The Need for Equity

Barry D. Stimmel, M.D.

Abstract

International Medical Graduates (IMGs) have played an integral role in provision of medical services in the United States, representing 25% of physicians nationwide and up to 43% in some states. Because of the concern with the future physician surplus, several proposals have been advanced specifically targeting IMGs. Although the need to downsize our residency training programs is apparent, focusing specifically on the IMG who has met all requirements for entry into residency training is neither morally acceptable nor pragmatically effective. Downsizing can be accomplished independent of the need to single out a specific group without jeopardizing either quality of training or quality of care if all groups responsible for residency training work together.

More than fifteen years have passed since a physician surplus was predicted by the year 2000 (1). Yet, despite little disagreement with this assessment, virtually nothing has been done by any agency or institution to diminish or even limit the number of physicians trained in this country. Between 1989 and 1993, the total number of residency positions in the United States increased by 19%, from 80,935 to 96,469 (2). Although it has also been agreed that we need more generalists or primary care physicians, if one looks at first-year positions (PGY-1) by specialty, the greatest increases were seen in those fields in which the greatest surpluses were predicted. Thus, between 1989 and 1994, primary care PGY-1 positions increased by only 3.4%, as compared with a 11.7% increase for the surgical specialties and a 30% increase for support specialties (3).

The response of organized medicine during this time has been directed to recommending a redistribution of the proportion of residency positions in primary care, so that these programs occupy 50% of all PGY-1 positions, and to committing medical schools to graduate 50% of their students to enter primary care fields. Yet, more recent data have suggested that, in fact, we already have too many primary care physicians, especially in view of the rapid advances of managed care into the medical marketplace (4, 5).

It has been predicted that by the year 2000, up to 65% of all Americans will be enrolled in a managed care network (6). Indeed, this may be a conservative estimate. Although controversy exists as to whether we have sufficient numbers of generalist positions, our present ratio of primary care physicians to 100,000 population in the United States of 88.8 is considerably more than that currently used by most health maintenance organizations or managed care groups (4–6). In these organizations primary care physician-to-population ratios approximate 49 physicians per 100,000 population, suggesting that, in fact, there are presently far too many physicians, even in primary care fields. Of course with respect to total physician-to-population ratios, our excesses,

---

Professor of Medicine and Medical Education, Dean for Graduate Medical Education, Mount Sinai School of Medicine of The City University of New York, New York, New York. Address requests for reprints to: Dr. Barry Stimmel, Dean for Graduate Medical Education, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1193, New York, NY 10029.

based on managed care data, are even greater (25 vs. 111 physicians per 100,000 population).

It should be apparent that if one wishes to downsize the number of physicians being trained, there are two obvious points of attack: the number of first-year positions in U.S. medical schools and/or the number of first-year positions in residency training. Recommendations have ranged from freezing the number of U.S. medical schools and their class sizes to closing 20% of U.S. medical schools (7). Medical schools have been quick to observe that, because the number of PGY-1 positions exceeds the number of U.S. graduates by more than 40%, diminishing the number of medical students trained in this country will have little or no effect on physician manpower but might well adversely effect the quality of care by filling existing residency positions with less well trained physicians. Indeed, parenthetically, despite the constant publicity given to physician oversupply, medical schools have been receiving the largest number of applicants in their history. The number of applicants applying for PGY-1 positions is more impressive, however. In 1995, there were 30,925 applicants to the National Resident Matching Plan (NRMP), of whom only 15,892 (51.4%) were U.S. citizens. These applicants competed for 20,751 positions (3).

Despite the acknowledged need to decrease residency positions, residency programs have been loath to diminish voluntarily their house staff. There have been several reasons advanced for this hesitancy. First, it has been correctly observed that in many hospitals, especially those serving medically indigent communities, residents are the only source of medical care available. Diminishing the numbers of house staff without appropriate replacement with other health care personnel would place a large number of communities at risk. Second, because residents are also correctly considered a "cheap source of labor," receiving a modest salary for their hours worked, employing them is extremely cost effective. Therefore, their elimination would be accompanied by a considerable rise in the cost of medical care. This second observation, in fact, is not precisely correct and perhaps is the main reason for the hesitancy of individual residency programs to cut house staff as well as for the increasing interest by Congress in downsizing residency training.

A major source of funding for graduate medical education comes from Medicaid and Medicare and is a considerable sum. This consists basically of direct medical education (DME) support, used to pay for house staff salaries and salaries of teaching faculty, and indirect medical education (IME) support, used to fund the added expenses needed to assist the teaching effort and for increased patient care costs incurred by teaching hospitals that care for severely ill uninsured persons. In 1994, Medicare DME and IME payments in the aggregate amounted to $5.6 billion dollars. In terms of individual resident reimbursement, in New York State, through Medicare and Medicaid payments, this figure came to approximately $200,000 per resident per year. Because residents' salaries and fringe benefits equal less than 25% of this amount, even with the reductions in payments proposed, on a per resident basis it still would not be cost effective for an institution to cut a single house staff position.

In the aggregate, however, savings to the government resulting from even small decreases in IME funding would be considerable. For example, the latest congressional proposal (H.R. 2425) to diminish IME funding from 7% to 6%, and finally, to 5.6% in fiscal year 2000, would save almost $6.7 billion over 7 years. Even more money could be saved, and the existing imbalance between physician supply and demand addressed, were a downsizing of all PGY-1 positions to occur. Unfortunately, as with most issues, politics frequently lords over reason. Despite recommendations toward this end by several groups, including those who produced the Clinton Health Plan, the recent proposed legislation freezes the number per hospital to that existing in July 1995 (8), rather than directly attacking the excess number of residency positions. In an attempt to still achieve considerable financial savings, the legislation seized on an approach which affects the International Medical Graduate (IMG). Focusing on the non-U.S. citizen IMG without a "green card," the proposal called for progressively diminished funding for those persons during the next 2 years, so that by 1998 non-U.S. citizen IMGs in residency training would receive only 75% of the funding provided to U.S. citizens. In essence, the legislation says to a hospital, if you want these residents, pay for them.

However, far from being without political influence, IMGs, through the International Association of American Physicians, have become a potent political force and, quite appropriately, are strongly supported by the American Medical Association and state medical societies (9). In addition, by focusing on those IMGs without a green card, the actual number of persons affected would be much less than expected because marriage remains the most common route to citizenship, with almost 50% of IMG residents in 1991 having a green card or being naturalized citizens (10). Not surprisingly, as this bill emerged from the House-Senate compromise, the section concerning the IMG was deleted. Subsequently, President Clinton refused to sign this legislation, resulting in the need to rethink how graduate medical education should be supported through Medicare funding.

It is not at all unlikely that the role of the IMG will once again become prominent and that the "easy fix" solution eliminating any reimbursement for IMGs will resurface. One assumes that the thinking behind this centers on the belief that if hospitals will not be reimbursed for house staff, they will not be hired. Not only is this incorrect reasoning but, in addition, it introduces the pernicious concept of promoting unequal pay for equal work based on citizenship or foreign medical school attended. To better define why such legislation will create more problems than it will solve, it is helpful to review the current status of IMGs in residency training in this country.

During the past 5 years the number of IMGs entering residency training has increased dramatically, from 5,780 in 1989 to 11,881 in 1995, with between 60 and 75% of all non-U.S. citizen IMGs completing their training remaining in this country to practice medicine (11, 12). Paradoxically, rather than making it more difficult for the non-U.S. citizen IMG to remain in this country, regulations have just been promulgated that actually increase the number of ways in which a physician may obtain a waiver of the need to return to his/her country of origin (13). The new regulations now allow a waiver to be obtained if any State or Department of Public Health, or its equivalent, petitions the Immigration and Naturalization Service (INS) to allow a physician to remain. In addition, the INS has dropped their proposal to prevent non-U.S. citizen IMGs from entering the United States under H-1B visas to pursue medical residencies. These visas, unlike the J1 visas, do not require physicians to return to their home country for 2 years before petitioning the INS for readmission to the United States.

Because the ultimate goal of all physicians, including IMGs, is to be licensed and to practice, the need to complete approved residency training becomes paramount over the need to earn income while engaging in this training. If proposed legislation does not address limiting the number of approved residencies but instead eliminates Medicare reimbursement for IMGs, they may well have to agree to work for less (or perhaps no) pay than U.S. citizens to complete the training required for licensure. Considering the need for hospitals to remain ever cost efficient, an incentive exists for hospitals to accede to such practices. Of course, not all hospitals will need to do this as those programs that are quite competitive and in the past have taken only highly qualified IMGs will now eliminate them from consideration because funding will no longer be provided. We will have thus fostered the development of a system by which selection will be based not on qualifications but on citizenship.

This scenario, although undesirable morally, is not even effective pragmatically because it does not address the need to downsize and coordinate residency training programs among institutions to prevent those smaller, less competitive institutions from being disproportionately affected in this process. Downsizing does not have to be accompanied by an adverse effect on patient care or a lessening of services provided to the indigent. It does require careful planning and the ability of institutions to share resources.

One solution, a variant of that recommended by the Council on Graduate Medical Education, as well as various other groups, is to develop consortia for graduate medical education consisting of a number of teaching hospitals. Each consortium would be charged with downsizing in the aggregate their residency positions to 110%–120% (or whatever percentage is considered appropriate) of U.S. medical graduates over a 5-year period. Funding for graduate medical education would be provided based on the number of residents in training in 1995, and held constant over a 7-year period to allow for the ability of the consortium to meet these goals without any adverse effects on provision of services. Because most hospitals are currently affiliated in some way with medical schools, the most logical approach would be for the consortium to be centered around the medical school. However, consortia could easily be formed around other medical centers as well. Under this system IME funding would continue to go to the individual hospital, whereas DME funding would flow to the consortium to allow for the establishment of a comprehensive and orderly system of house staff training. This would allow a cooperative rather than a competitive effort to develop, with each institution in the consortium being able to use each other's strengths. Models of well-functioning consortia already exist and could be replicated easily. This had been partially addressed in H.R. 2425 by establishing a system of IME reimbursement based on the average of previous payments between 1992 and 1994, independent of current or future resident size. The establishment of a trust fund for graduate medical education to assist teaching hospitals in supporting the added costs of educating house staff was also quite positive. However, unless associated with shrinking the number of residents in training, the increasing oversupply of physicians will persist.

The process of downsizing may not be as difficult as one might think. Downsizing of PGY-1 positions would be immediately facilitated by eliminating all of the preliminary PGY-1 positions in medicine and surgery that are not directly linked to advanced residency training. Decreasing available positions in specialties that currently are unable to match despite the plethora of candidates, such as anesthesiology and radiology, could also occur quickly. Market forces are already in play influencing the number of U.S. students entering primary care fields without the need to either coerce any student or penalize any medical school for those students who choose not to become generalists. The Accreditation Council for Graduate Medical Education (ACGME), by taking an aggressive attitude in disapproving marginal programs, will further serve to downsize the residency base to a more suitable number. Finally, by restructuring waivers the number of non-U.S. citizen IMGs remaining after finishing training would also decrease. This, in turn, would facilitate the return of highly trained physicians to their own countries, many of which are severely medically underserved. For this to occur, however, behavioral changes are needed. Individual institutions would have to be willing to relinquish complete autonomy, the ACGME would have to be willing to take a much firmer stand on marginal programs than it has in the past, and funding for GME would have to be separated from the "per resident" formula. This last change would serve as a catalyst for the others by providing a financial incentive for initiating the downsizing process.

Downsizing the resident work force would result in "uncovered beds" and the need to replace residents with other health care providers. The use of these health care providers has been criticized as too expensive (a 3:1 ratio is suggested) and impractical because they are too few in number to meet the needs if a 30% reduction in residencies were to occur. However, in view of existing market forces, this may be far less of a problem. First, the use of hospital beds is diminishing markedly because of a shift to the ambulatory setting. Second, the availability of board-eligible physicians is becoming noticeable as managed care moves into communities. Staff physicians can see patients much more efficiently than residents, can work nights and weekends, order fewer expensive procedures, and can bill for services. Attending-directed services can be developed, which, when combined with loan forgiveness after it is established in underserved areas, may be quite attractive to a number of persons completing training.

Whitcomb has recently suggested that downsizing be addressed by limiting the number of graduates of foreign medical schools who would be allowed to enter residency training, rather than regulating the number of residency positions (5). Although this is a rational approach to the issue of the IMG, it nonetheless is less than satisfactory. To be eligible to enter residency training, all IMGs must pass the same examinations as U.S. medical graduates and are therefore equally qualified. By limiting the number of residency positions, an open system will be created whereby each program director can make independent choices. By focusing on the number of PGY-1 positions, rather than the citizenship of those who fill them, the issue of equity with respect to the IMG would be resolved.

Our country has always been one that has welcomed people from other countries, allowing them to compete in the marketplace based on their ability. Although no one should be guaranteed a profession, no one who meets uniform standards should be summarily discriminated against because of federal guidelines. There is no reason to deviate from this policy because our goals could be met better by its preservation. In addition, it is not the right thing to do.

References

1. Summary Report of the Graduate Medical Education Advisory Committee, vol. 1. Washington, DC: Department of Health and Human Services, 1980.
2. Graduate Medical Education. J Am Med Assoc 1994; 272: 729–730.
3. NRMP Data, March 1995. Washington DC: National Resident Matching Program, 1995.
4. Weiner JP. Forecasting the effects of health reform on US physician work force requirement: evidence from HMO staffing patterns. J Am Med Assoc 1994; 272:222–230.
5. Whitcomb M. A cross national comparison of generalist work force data. Evidence for US supply adequacy. J Am Med Assoc 1995; 274:692–695.
6. Sixth Report of the New York State Council on Graduate Medical Education 1994, New York, NY.
7. Health panel urges U.S. to cut funding for training of foreign born physicians. Washington Post January 24, 1996: A1:7.
8. Subtitle E: Reform of payment for graduate medical education and teaching hospital payment. Washington DC:

Congress of the United States, House of Representatives, September 1995.

9. Page, L. Once seeking separation, IMGs now swimming in medical mainstream. Am Med News February 28, 1994: 3,7.
10. Page, L. Filling the gaps. Am Med News February 21, 1994: 7.
11. Mullan F, Politzer RM, Davis CH. Medical migration and the physician work force: international medical graduates and American medicine. JAMA 1995; 273:1521–1527.
12. GME Census. Washington, DC: Association of American Medical Colleges SAIMS Database, 1995.
13. COTH Council of Teaching Hospitals Selected Activities Report. Washington DC: Association of American Medical Colleges, May 1995.

Isoniazid Hepatotoxicity after Orthotopic Liver Transplantation

Leona Kim Schluger, M.D., Patricia A. Sheiner, M.D., Mark Jonas, M.D., James V. Guarrera,
Isabel M. Fiel, M.D., Burt Meyers, M.D., and Paul D. Berk, M.D.

Abstract

The hepatotoxic effects of isoniazid have been well described, but there have been no reports on the incidence of isoniazid-induced liver disease in patients who have received an orthotopic liver transplant. We retrospectively reviewed the records of 13 patients who received isoniazid after liver transplantation for either chemoprophylaxis or as part of a multidrug regimen for the treatment of *Mycobacterium tuberculosis* infection. Five of the 13 patients developed biochemical and histologic evidence of isoniazid hepatotoxicity. All five patients were on a multidrug regimen which included the administration of rifampin. No hepatotoxicity occurred in patients who received isoniazid alone or in conjunction with ethambutol for chemoprophylaxis. In conclusion, the incidence of isoniazid hepatotoxicity increased when the drug was used in conjunction with rifampin for the treatment of *M. tuberculosis* infection.

Isoniazid (INH) was introduced in 1952 for the treatment of tuberculosis. The hepatotoxic effects of INH were not appreciated until two decades after its introduction, when several fatalities caused by the medication were documented (1). The reported incidence of INH hepatitis varies with age and ranges from 0 to 20% (2–4), but there have been no reports on the incidence of INH hepatitis among liver transplant recipients. We retrospectively reviewed our experience with liver recipients at The Mount Sinai Medical Center who received INH either for treatment of documented tuberculosis or for chemoprophylaxis to determine whether the incidence of INH hepatitis is greater in liver recipients and to determine whether any risk factors for toxicity could be identified.

Patients and Methods

We reviewed our data on all adults who underwent orthotopic liver transplantation at The Mount Sinai Medical Center between September 1988 and December 1995 and who subsequently were placed on INH for treatment of tuberculosis or for chemoprophylaxis.

Diagnosis in liver recipients with clinical signs and symptoms of tuberculosis was confirmed by positive acid-fast bacillus (AFB) stains of sputum samples or tissue biopsies and/or AFB cultures. Patients with biopsy- or culture-proven *Mycobacterium tuberculosis* infection were placed on a four-drug regimen of isoniazid (300 mg/day), rifampin (RMP), (600 mg/day), ethambutol (EMB) (15 mg/kg/day), and pyrazinamide (PZA) (20 mg/kg/day), or a five-drug regimen with ofloxacin (400 mg twice daily), for up to 1 year.

Patients who had positive tests for purified protein derivative of tuberculin during pretransplant evaluation and abnormal baseline chest x-rays, but no evidence of active tuberculosis, were placed on INH chemoprophylaxis after liver transplantation. These patients, as well as patients who had had close contact with an individ-

---

From the Departments of Medicine (LKS, MJ, BM, PDB), Surgery (PS), and Pathology (IMF), and The Mount Sinai School of Medicine (JVG), The Mount Sinai Medical Center, New York, New York. Address reprint requests to Leona Kim Schluger, M.D., Mount Sinai Hospital, Box 1104, One Gustave L. Levy Place, New York, NY 10029.

TABLE 1
Patient Demographics

Patient No.	Age (yr)	Sex	Diagnosis[a]	Donor Age (yr)	Tuberculosis Medications	LFTs at INH Initiation		
						BILI (mg/dL)	alk phos	ALT (U/L)
1	41	M	HCV	34	5[e]	2.3	1131	24
2	29	M	Wilson	23	4[d]	0.6	66	29
3	27	M	PSC	18	4[d]	0.4	249	32
4	49	F	HCV	54	4[d]	1.8	405	89
5	54	M	HCV/HBV	61	4[d]	1.8	241	53
6	27	M	Budd-Chiari	21	INH EMB	0.3	69	53
7	67	M	EtOH	31	INH	1.4	206	22
8	50	M	EtOH	15	INH EMB	1.0	84	9
9	53	M	HCV	39	INH	0.7	84	34
10	57	M	HCV	50	INH	1.1	184	15
11	61	F	HCV	40	INH	1.4	92	92
12	49	M	PBC	64	INH	1.9	131	442
13	24	M	Budd-Chiari	18	INH	1.2	90	14

[a] HCV, hepatitis C virus; PSC, primary sclerosing cholangitis; HBV, hepatitis B virus; PBC, primary biliary cirrhosis.
[b] FK, FK506 (tacrolimus); CYA, cyclosporine.
[c] Rej, rejected; hep, hepatitis; bx.
[d] INH, RMP, EMB, PZA.
[e] INH, RMP, EMB, PZA, Ofloxacin.

ual who had a recent infection with *M. tuberculosis,* received chemoprophylaxis with INH alone (300 mg/day) or INH and EMB (15 mg/kg/day). In all patients, INH therapy was discontinued upon confirmation of INH-induced hepatotoxicity.

We recorded indication for INH therapy, age, donor age, duration of INH therapy, reason for discontinuation of INH, immunosuppressive regimen, rejection episodes during INH therapy, other concurrent antituberculosis drug therapy, and the administration of other medications. We also recorded levels of total bilirubin (TBIL), alkaline phosphatase (alk phos), alanine transaminase (ALT), and aspartate transaminase (AST) at the start of INH therapy, when abnormal liver function tests (LFTs) prompted a liver biopsy, and at termination of therapy. Liver biopsy results were also noted.

Liver Pathology

Liver biopsy specimens were obtained when clinically indicated by elevations in liver chemistries. Five-micrometer-thick sections of formalin-fixed, paraffin-embedded core biopsy specimens were stained with hematoxylin and eosin (H&E). Histologic features consistent with acute hepatitis were defined by the presence of lobular activity, particularly necroinflammatory lesions and acidophilic bodies, along with activation of sinusoidal lining cells. The diagnosis of isoniazid toxicity as the cause of the acute hepatitis was at-

TABLE 1
Continued

Patient No.	LFTs at INH Discontinuation			Time on INH (months)	Immunosuppression[b]	Biopsies during INH Therapy[c]	Other Medications[d] (in addition to septra, acyclovir, mycelex)
	BILI (mg/dL)	alk phos	ALT (U/L)				
1	16	913	131	3	FK	1. Rej 2. INH	nifedipine, famotidine
2	0.5	94	688	2	FK	1. Rej 2. Rej + INH	furosemide
3	5.4	252	378	4	CYA	1. INH	azulfidine, famotidine
4	3.2	163	214	1	CYA	1. INH + HCV	furosemide, famotidine, nifedipine
5	8.6	547	201	1	CYA	1. INH	
6	0.7	71	19	15	FK	no bx	warfarin, nifedipine
7	2.3	142	35	6	CYA	1. Rej 2. Rej 3. Rej	colchicine, epoetin alfa
8	0.8	64	116	16	CYA	1. Rej 2. ETOH	furosemide, allopurinol, colchicine
9	1.1	118	12	31	CYA	no bx	nifedipine enalapril
10	1.1	128	298	6	CYA	1. Rej 2. HCV	nifedipine, famotidine
11	0.9	62	40	9	CYA	no bx	diltiazem, famotidine
12	0.1	91	18	14	CYA	1. Rej	nifedipine, levothyroxine
13	0.6	86	35	11	CYA	no bx	nifedipine, furosemide, warfarin

tained by the exclusion of other causes of hepatitis, particularly viral and steatohepatitis, and/or acute cellular rejection (5).

Results

Of 918 adults who underwent orthotopic liver transplantation at our institution between September 1988 and December 1995, 14 (1.53%) were placed on INH. One patient died 4 months post-transplant because of recurrent hepatitis B virus and was excluded from this analysis. Five of the remaining 13 patients (38.5%) received multidrug therapy for documented active tuberculosis. In two of the five patients, the diagnosis was confirmed with positive AFB stain and culture from pleural biopsies; in the other three patients, diagnosis was confirmed with a lung biopsy, a mediastinal lymph node biopsy, and cerebrospinal fluid, respectively. Eight patients (61.5%) received INH alone or INH and EMB for chemoprophylaxis.

Demographic data are summarized in **Table 1**. The mean age at orthotopic liver transplantation was 45 ± 13.7 years (range, 24–67 years). Mean donor age was 36 ± 13.0 (range, 14–64 years). Mean LFTs at the start of INH therapy were ALT, 70 ± 110 U/L; AST, 50 ± 24 U/L; TBIL, 1.8 ± 0.6 mg/dL; alk phos, 233.2 ± 275.9 U/L.

In all five patients receiving INH as part of the multidrug regimen for documented active tuberculosis, liver biopsy performed to evaluate worsening LFTs revealed evidence of INH hepatoxicity, and INH therapy was discontinued. In these five patients, mean LFTs at discontinuation of INH therapy were: ALT, 322 ± 200 U/L; AST, 266 ± 190 U/L; TBIL, 6.7 ± 5.3 mg/dL; and alk phos, 394 ± 302 U/L. Three of the five patients presented with jaundice; all were otherwise clinically asymptomatic. The interval between initiation of INH therapy and the diagnosis of INH hepatitis ranged from 1 to 4 months (mean, 2.4 months).

Biopsy findings in patients with INH-associated hepatitis are listed. Liver biopsy obtained in patient 1 after 3 months of therapy showed acute INH hepatitis with necroinflammatory lesions in the lobules and acidophilic bodies and activation of sinusoidal lining cells (Fig. 1a). Bridging necrosis connecting adjacent centrilobular areas was also seen occasionally with resultant dropout of hepatocytes (Fig. 1b). There was no evidence of acute cellular rejection or viral hepatitis. These features are characteristic of INH hepatitis. In patient 2, liver biopsy performed at 2 months showed necroinflammatory lesions in the parenchyma and a mixed cellular infiltrate in the portal tracts, endothelialitis, and bile duct damage consistent with mild to moderate cellular rejection (Fig. 2a) in addition to acute drug-induced hepatitis (Fig. 2b). In patient 3, biopsy findings at 4 months were consistent with INH hepatitis without evidence of rejection or viral hepatitis. In patient 4, biopsy obtained at 1 month showed a combined picture of drug-induced hepatitis and recurrent hepatitis C virus infection. Patient 5 exhibited acute INH hepatitis in a biopsy in his first month of treatment.

None of the eight patients who received INH for chemoprophylaxis developed INH-associated hepatitis ($p < 0.001$). The duration of INH therapy in this group ranged from 6 to 31 months (mean, 13.5 months). Four of these patients did develop transient worsening of liver chemistries (Table 1) during therapy which prompted the performance of liver biopsies. One patient had three episodes of acute cellular rejection, the second had one episode of rejection and one biopsy consistent with acute alcoholic hepatitis, the third had one biopsy consistent with acute cellular re-

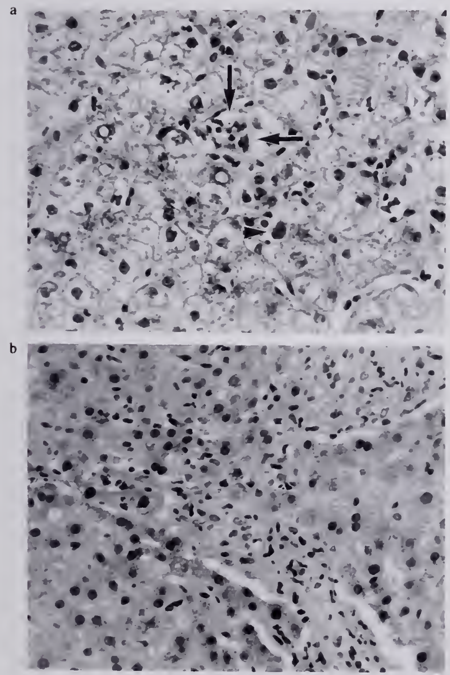


Fig. 1. (a) Foci of parenchymal necrosis (arrows) and acidophilic bodies (arrow heads). Some hepatocytes are undergoing ballooning degeneration. (H&E; original magnification ×40). (b) Presence of bridging necrosis (H&E; original magnification ×40).

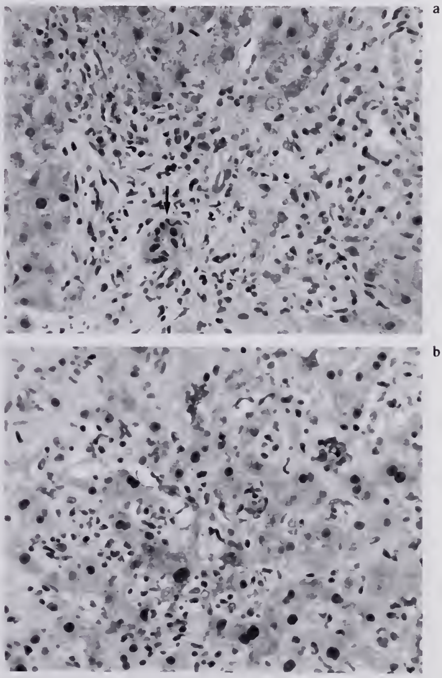


Fig. 2. (a) Portal tract showing typical histologic features of acute cellular rejection consisting of a mixed inflammatory cell infiltrate and bile duct damage (arrow) (H&E; original magnification ×40). (b) In the same biopsy (a) diffuse necroinflammatory lesions and acidophilic bodies are found (H&E; original magnification ×40).

jection and recurrent hepatitis C infection, and the fourth had one episode of acute cellular rejection.

Development of INH hepatitis was strongly correlated with administration of the multidrug regimen ($p < 0.001$). There was no correlation between development of INH hepatitis and LFTs at initiation of therapy, incidence of rejection, donor age, or type of immunosuppressive regimen.

Discussion

Even as the twentieth century draws to a close, tuberculosis remains a diagnostic and therapeutic challenge for the clinician. In the field of transplantation, the development of more potent immunosuppressive agents has coincided with the emergence of an increasing number of infectious complications among transplant recipients. Tuberculosis is a particular problem in liver transplant recipients, because the treatment involves use of known hepatotoxic medications.

Streptomycin, introduced a half century ago, was the first antibiotic available against tuberculosis. Isoniazid, a bactericidal antibiotic, was developed in 1951; the first clinical trial was conducted by Robitzek and Selikoff (6) at the Seaview Sanatorium in Staten Island, New York. Initial reports on the efficacy and safety of INH were encouraging. Trials of INH conducted by the U.S. Public Health Service (USPHS) between 1955 and 1968 resulted in very few cases of hepatotoxicity (7).

Only in the early 1970s were the potentially lethal effects of INH appreciated, when an unexpected clustering of hepatitis was observed among individuals taking INH in a tuberculosis control program. In 1972, Garibaldi et al. (8) reported on 19 cases of clinical hepatitis among 2321 patients (1%) receiving INH chemoprophylaxis. Thirteen of the 19 patients developed jaundice and two died (0.1%). These observations by Garibaldi and co-workers prompted a multicenter study by the USPHS. In 1978, the USPHS reported on 13,838 patients receiving INH prophylaxis: 1% developed hepatitis (10.3 per 1000) and 8 (0.06%) died (7).

A review of the literature in 1992 by Israel et al. (3) revealed that the reported incidence of hepatitis in patients receiving INH prophylaxis ranged from 0.3 to 6.4%, with a hepatitis case fatality rate of 0.0 to 8.7%. Factors predisposing patients to development of INH-related liver damage included alcohol use and age greater than 35 years. The incidence of INH hepatotoxicity was 0% for persons younger than 20 years, 0.3% for ages 20 to 35 years, 1.2% for ages 35 to 49 years, and 2.3% for ages greater than 50 years (9).

The pathogenesis of INH hepatotoxicity is unknown. The mechanisms that have been postulated include a hypersensitivity reaction and the formation in the liver of a hepatotoxic metabolite (10). Maddrey and Boitnott (11) reported cases of an allergic-type reaction to INH manifested by fever, chills, and worsening liver functions a few hours after rechallenge with isoniazid. However, other studies have failed to show other signs and symptoms of an allergic reaction such as skin rash and peripheral eosinophilia (10). Additionally, the onset of liver injury from INH usually occurs after prolonged exposure rather than after the characteristic fixed period (1–4 weeks) of exposure associated with an allergic mechanism of injury (10).

The formation of a hepatotoxic metabolite through acetylation of INH was proposed by Mitchell and Jollow (12) as the causative mechanism in INH-induced liver damage. The authors observed that "rapid acetylators" were more likely to develop hepatic injury than "slow acetylators." Clinical observations and experimental data appear to support strongly the hypothesis that INH-induced liver injury is caused by the formation of a hepatotoxic metabolite (10).

In addition to the association of increasing age and alcohol consumption with the development of INH-related liver damage, the addition of RMP to INH for the treatment of tuberculosis has been shown to act synergistically in producing toxic hepatitis. The pooled data from the literature reviewed by Steele et al. (13) in 1991 showed the incidence of hepatotoxicity to be 2.55% in patients taking INH plus RMP as compared with 1.6% in patients given INH without RMP ($p = 0.048$). The incidence of hepatitis in the group which received INH plus RMP was also higher than that in the group which received RMP without INH (1.1%) ($p = 0.008$). One of the proposed mechanisms for the increased toxicity with the concurrent use of INH and RMP is the increased production of toxic metabolites of INH from the potent inducing effect of RMP on the hepatic P-450 enzyme pathway.

Aside from INH and RMP, pyrazinamide has also been shown to cause liver injury. The reported hepatotoxicity of pyrazinamide, however, occurs only when the drug is given in large doses and for a prolonged period (14). A short course of PZA (25–35 mg/kg for 2 months) produced no increase in the incidence of hepatotoxicity (15).

There have been no reports on the incidence of INH-induced hepatitis in liver transplant recipients. In our small series, the overall incidence of INH hepatitis among the 13 patients who received the drug was 38% (5 of 13 patients). It is important to emphasize that all five patients who had liver biopsies and liver chemistries compatible with drug-induced hepatitis were on a multi-antituberculosis drug regimen that included administration of rifampin. None of the eight patients who received INH alone or in combination with EMB for chemoprophylaxis developed clinical hepatitis. Although PZA has been shown to cause hepatotoxicity and was used as part of the multidrug regimen in our population, hepatotoxicity is rare at the dosage administered to our patients. The observations made in this study suggest that the liver toxicity of INH is compounded in particular by the concommitant administration of rifampin.

There were no deaths in our study group. It should be noted that our transplant recipients are observed closely, with monthly laboratory tests, regardless of whether they are receiving antituberculosis medications, and any abnormalities in liver chemistries are promptly investigated. In the five patients who had evidence of INH-induced hepatitis, all potential hepatotoxic medications (INH, RMP, and PZA) were discontinued with subsequent improvements in their liver functions. Thereafter, the patients were maintained on EMB and ofloxacin for the remainder of the treatment period. The cautious clinical follow-up of our liver transplant recipients, especially those who are receiving potentially hepatotoxic medications, most likely accounts for the favorable outcome of the study patients who developed evidence of INH-induced liver injury.

Conclusions

Among patients who have undergone liver transplantation, there seems to be no added risk for development of INH-induced hepatotoxicity when the drug is administered for chemoprophylaxis. However, there seems to be a significant increase in INH-associated liver disease among patients treated with a multidrug regimen that includes RMP. Given this observation of an increased incidence of INH hepatotoxicity when used in combination with rifampin, alternative regimens are currently under consideration for the treatment of *M. tuberculosis* infection among the liver transplant recipients.

References

1. Salpeter SR. Fatal isoniazid-induced hepatitis. Its risk during chemoprophylaxis. West J Med 1993; 560–564.
2. Farrell FJ, Keeffee EB, Man KM, et al. Treatment of hepatic failure secondary to isoniazid hepatitis with liver transplantation. Dig Dis Sci 1994; 39:2255–2259.
3. Israel HL, Gottlieb JE, Maddrey WC. Perspective: preventive isoniazid therapy and the liver. Chest 1992; 101: 1298–1301.
4. American Thoracic Society. Treatment of tuberculosis and other mycobacterial diseases. Am Rev Respir Dis 1983; 127:790–796.
5. Thung SN, Gerber MA. Acute hepatitis versus cellular (acute) liver allograft rejection. In: Thung SN, Gerber MA, editors. Differential diagnosis in pathology: liver disorders. New York/Tokyo: Igaku-Shoin, 1995: 920–932.
6. Robitzek EH, Selikoff IJ. Hydrazine derivatives of isonicotinic acid (Rimifon, Marsilid) in the treatment of active progressive caseous-pneumonic tuberculosis. A preliminary report. Am Rev Tuberc Pulm Dis 1952; 65: 402–428.
7. Kopanoff DE, Snider DE, Caras GJ. Isoniazid-related hepatitis. A U.S. public health service cooperation surveillance study. Am Rev Respir Dis 1978; 117:991–1001.
8. Gabribaldi RA, Drusin RF, Ferebee SH, et al. Isoniazid-associated hepatitis: Report of an outbreak. Am Rev Respir Dis 1972; 106:357–365.
9. Centers for Disease Control. Isoniazid associated hepatitis: summary of the report of the Tuberculosis Advisory Committee and special consultants to the Director. MMWR 1974; 23:97–98.
10. Black M, Mitchell JR, Zimmerman HJ, et al. Isoniazid-associated hepatitis in 114 patients. Gastroenterology 1975; 69:289–302.
11. Maddrey WC, Boitnott JK. Isoniazid hepatitis. Ann Intern Med 1973; 70:1–12.
12. Mitchell JR, Jollow DJ. Metabolic activation of drugs to toxic substances. Gastroenterology 1975; 68:392–410.
13. Steele MA, Burk RF, DesPrez RM. Toxic hepatitis with isoniazid and rifampin: A meta-analysis: Chest 1991; 99:465–471.
14. Matthews J. Pyrazinamide and isoniazid used in the treatment of pulmonary tuberculosis. Am Rev Respir Dis 1960; 81:348–351.
15. Fox W, Mitchison DA. Short course chemotherapy for pulmonary tuberculosis. Am Rev Respir Dis 1975; 111: 325–353.

Thomas C. Chalmers In Memoriam

Dr. Thomas C. Chalmers

I FIRST MET TOM CHALMERS 40 years ago when we were among the founding members of the first clinical trials group to study cancer therapy, the Eastern Solid Tumor Group subsequently named the Eastern Cooperative Oncology Group (ECOG). He had recently completed his landmark study on hepatitis at Walter Reed Army Medical Center, a classic investigation that took more time to plan than it did to execute. The study had a faultless design, meticulous execution, and believable conclusions. The *Journal of Clinical Investigation* devoted an entire separatum to this study, the longest paper they ever published. Contrary to conventional clinical wisdom, and to army practice, Tom found that strict bedrest was neither essential nor beneficial for soldiers with acute hepatitis. Although his career might have been influenced by this study, I think it was the other way around. Conventional clinical wisdom never recovered.

He approached every controversial clinical problem with a passion, indeed a need to determine the true answer by the powerful tool, the randomized controlled trial. He headed the professional services of the Lemuel Shattuck Hospital, a Massachusetts state facility for chronic care. From that vantage point he founded the Boston Liver Group and conducted numerous controlled trials evaluating therapeutic approaches to cirrhosis, such as prophylactic portocaval shunt, which was shown not to be useful.

When I first considered Mount Sinai as a new base, I was interviewed by George James, the first President. Shortly thereafter, he died suddenly. Hans Popper, the founding father of the Mount Sinai School of Medicine became the acting president, and upon learning that Tom and I were friends, asked me to add to the chorus of those urging Tom to come to Mount Sinai to assume the presidency. I visited Tom and Frankie in Bethesda, where he was the Director of the Clinical Center of the National Institutes of Health, and I like to think I nudged him in the right direction.

Some presidents of organizations such as this one are concerned with, remembered for, and memorialized in hardware. They are the builders, the bricks men. Tom was a software president. He provided the intellectual mortar that propelled Mount Sinai School of Medicine to excel. His conditions for coming to Mount Sinai were two. First, he required that the trustees allow him to establish a Department of Biomathematical Sciences. He saw this as a mechanism to blend science with clinical medicine, to study diagnostic and therapeutic interventions by well-designed controlled trials that would measure the effects and assess their significance. His second requirement was to revitalize the Department of Continuous Education, because he knew that 4 years of medical schooling at the outset of a 40-year career were merely table stakes anteed up to get in the game. The frequent innovations in medicine, new, but not necessarily better—that's where the randomized trials would count—were, without a strong commitment, institutional and personal, to lifelong education, a guarantee of anachronistic medicine and obsolescent doctors.

Tom will be remembered for being an educator and an experimentalist. "Teach our students the love of learning, the scientific method, the critical analysis of the literature and they will continue to be good doctors." He was an experimentalist in every aspect of his life. His driver would ask which road to Bronxville—and Tom, still trying to prove that one route was faster, would flip his coin for a random choice. He was stymied, however, when, deploring the unpredictable performance of medical students despite a grueling admissions process, he suggested that a segment of the first-year class be chosen at random from the eligibles—opposed of course by those who had a son or daughter who might want to follow the family tradition of medicine at Mount Sinai.

During his presidency, Tom was highly magnetic. He attracted medical students to work with him analyzing the literature about controversial topics. With exhaustive library searches and Chalmersian standards for the reports they found, Tom and his British friend Richard Peto were the prime exponents of the methodology of evaluation called meta-analysis. Using this method, he addressed topics as diverse as adjuvant therapy for breast cancer—useful—and oral antibiotic administration prior to gastrointestinal surgery—so well proved that further administration of placebo was unethical. It is not surprising that he was selected as a trustee of the National Library of Medicine.

Educator, experimentalist, and friend, to his faculty, his colleagues, and his students, Tom's was not a lofty perch. He fraternized with all, not because of noblesse oblige—he was one of us. He was a man of utmost integrity. A pledge made was a promise kept.

I had occasion to extol Tom in life more than a decade ago in somewhat mechanical terms. "His understanding of the indissoluble union between today's science and tomorrow's medicine has inspired us from the front, propelled us from the back and supported us from below, but never suppressed us from above."

Tom nurtured the academic esprit that electrified this nascent school in the 70s and 80s. His is a legacy of excellence in medicine to which he pointed us all, and which he himself achieved. Mount Sinai was changed by his presidency, and much for the better.

James F. Holland, M.D.
Mount Sinai Medical Center

Perturbing Gene Expression with Oligodeoxynucleotides:

Research and Potential Therapeutic Applications

Alan M. Gewirtz, M.D.

Abstract

The ability to block gene function with antisense oligodeoxynucleotides has become an important tool in many research laboratories. Because activation and aberrant expression of proto-oncogenes seems to be an important mechanism in malignant transformation, targeted disruption of these genes and other molecular targets with oligodeoxynucleotides could have significant therapeutic utility as well. In this regard, the potential therapeutic usefulness of oligodeoxynucleotides has been demonstrated in many systems and against several different targets including viruses, oncogenes, proto-oncogenes, and an increasing array of cellular genes. These studies in aggregate suggest that synthetic oligomers have the potential to become an important new therapeutic agent for the treatment of cancer in humans. Nevertheless, it is clear that considerable optimization will be required before antisense oligonucleotides will emerge as an effective agent for treating disease in humans. Progress will need to occur on several fronts. Included are issues related to the chemistry of the molecules used: for example, how chemical modification has an impact on uptake, stability, and hybridization efficiency of the synthetic DNA molecule. A clearer understanding of the mechanism of antisense-mediated inhibition, including where such inhibition takes place, will also be required. Finally, cellular "defense" mechanisms, such as increasing transcription of the targeted message, may also be factors to consider in planning effective treatment strategies with these agents. Finally, choice of target is also obviously an important issue.

THE PAST 15 YEARS have been a "golden age" in the field of hematology. During this time many of the extracellular growth factors that regulate hematopoietic cell development have been molecularly cloned, expressed as active proteins, and have been used in the clinic where they appear to be making a genuine contribution to patient management. In addition, the intracellular events that are triggered when growth factors interact with their cognate receptors on hematopoietic cells have also begun to be defined. In the past few years, in particular, several important elements of the signal transduction pathways used and transcription factors activated by the signaling apparatuses have been identified. Nevertheless, most of the molecular machinery that regulates blood cell development remains hidden in the "black box" which lies below the cell surface. This machinery, in normal blood cells in particular, has been largely inaccessible because of the difficulty of applying modern molecular analytic techniques to the small numbers of cells which are ordinarily available for such investigations.

Two main strategies have been applied to understanding gene function in hematopoietic cells.

This work was supported by grants from the R01 CA 54384, CA 66731, and PO1 CA 51083 from the National Institutes of Health.

Professor of Pathology and Internal Medicine, University of Pennsylvania School of Medicine.

Adapted and updated from the author's presentation at the Mount Sinai Medical Center on January 14, 1993. Address reprint requests to Alan M. Gewirtz, M.D., Room 513B, Stellar-Chance Laboratories, University of Pennsylvania School of Medicine, 422 Curie Blvd., Philadelphia, PA 19104.

One approach involves infecting cells with a vector engineered to express the gene of interest at high levels. If the cell's phenotype/behavior changes in the infected cell, but not a sham-transfected cell, one can tentatively attribute the change in phenotype to the newly expressed gene. Function may therefore be inferred. This approach, although potentially informative, has several drawbacks. First, it is not certain that the changes observed are directly related to the gene's function. Any effect observed may be an indirect one. Second, the experiment is not really "physiologic" because overexpression of the gene may exaggerate its normal function or impart new functions by leading to an excess of the encoded protein. Finally, this approach is often limited to leukemic cell lines because normal cells are often much more problematic to work with in terms of both infection and expression of the vector. An alternative to the overexpression strategy is to either physically "knock out" the gene of interest or to interfere with its function by perturbing the use of the gene's encoded mRNA. This approach has the theoretical advantage of being more predictive of function under physiologic conditions because any resulting phenotype is presumably the direct result of the loss of function of the targeted gene. It must be clearly stated nonetheless that any phenotype observed may affect the final result of events downstream of the targeted perturbation. In addition, these of "loss-of-function" experiments may be carried out more easily in normal cells. For these reasons, this approach has become an increasingly popular strategy for exploring gene function in cells of virtually any type.

Disrupting the function of a gene, or genes, which are thought to play a primary role in the pathogenesis of a given disease has obvious therapeutic appeal. Such an approach could allow for the development of disease-specific therapies which might be carried out with little or no normal tissue toxicity. Once consigned to the realm of science fiction, the technological means for accomplishing this goal are now at hand. In the laboratory, methodologies used for the perturbation of specific genes include homologous recombination, triple helix formation, and the use of reverse complementary DNA or RNA to interfere with utilization of the target gene's mRNA. The latter are the so-called "antisense strategies." Each of these methods for manipulating gene expression relies on nucleotide base pairing for specifically targeting the gene of interest. Each also has technical limitations which compromise its utility. For example, homologous recombination occurs only in dividing cells and is itself a very rare event. Triplex formation depends on the formation of Hoogsten bonds which require long spans of homopurine/pyrimidine bases for stability. Finding an appropriate sequence can therefore be problematic. Antisense strategies have their own limitations relating to uptake and stability. In addition problems related to the physical oligodeoxynucleotide (ODN)-target mRNA interactions are also difficult to predict. Cellular mRNAs are thought have highly ordered secondary and tertiary structures. If the targeted sequence is buried within the core of a structurally complicated mRNA molecule, or rendered inaccessible by an associated protein, then hybridization of the ODN and mRNA may not be possible. Generally, it is difficult to predict the secondary structure of mRNA molecules, and accessible regions must often be determined empirically by trial and error. Nevertheless, compared with the other available techniques, the antisense approach is relatively simple and straightforward. This is especially true for antisense oligodeoxynucleotides (AS ODN) which do not rely on viral vectors for introduction into cells.

What I would like to tell you about now is work that we have conducted in the course of targeting to target two different proto-oncogenes: the c-*myb* proto-oncogene, which encodes a nuclear transcription factor, Myb, and the c-*kit* receptor, a member of the type III tyrosine kinase family of cell surface receptors.

The Proto-oncogene c-*myb*

Of the genes that we have targeted for disruption using an AS ODN strategy, one that has been of particular scientific interest in our laboratory, and one in which therapeutically motivated disruptions are now in clinical trial, is the c-*myb* gene. The proto-oncogene c-*myb* is the normal cellular homologue of v-*myb*, the transforming oncogene of the avian myeloblastosis virus and avian leukemia virus E26. It is a member of a family composed of at least two other highly homologous genes designated A-*myb* and B-*myb*.

Located on chromosome 6q in humans, the predominant transcript of c-*myb* encodes a ~72 kDa protein called Myb. Myb protein consists of three primary functional regions. At the NH_2 terminus is a DNA binding domain which recognizes the core consensus sequence 5′-pyAAC(G/Py)G-3′. This region consists of three imperfect repeats consisting of 51-52 amino acids which have been designated R1, R2, and R3, respectively. R2 and

R3 have been shown to be absolutely required for DNA binding, whereas the role of R1 is less clear. It may contribute to the stability of the DNA-protein interaction but it does not appear to be required for binding. Within each repeat are three perfectly conserved tryptophan residues. Together they form a cluster in the hydrophobic core of the protein which maintains the DNA binding helix-turn-helix structure.

The midportion of the Myb protein contains an acidic transcriptional activating domain. In concert with the DNA-binding portion of the protein, these two regions are sufficient to transactivate promoters containing Myb binding sites. An immediately adjacent downstream region of the protein also appears to play a role in the activation of some targets, e.g., c-*myc*. This same region, independent of the DNA binding domain, has also been reported to be capable of transactivating certain targets such as the heat shock protein and DNA polymerase α. Finally, it has also been reported that some Myb transactivation targets require an interaction between Myb and other transcription factors, such as the nuclear factor-m (NF-M), and the CAT enhancer-binding protein (C/EBP). It is therefore likely that some of Myb's transcriptional activating properties may depend on the protein partners with which it interacts. This of course appears to be true of many transcriptional factors.

Myb also contains a negative regulatory domain which has been localized to the carboxy terminus. Interestingly, the carboxy terminus is deleted in v-*myb;* this has been thought to contribute to v-*myb's* transforming ability. Recently, a putative leucine zipper structure was described within the amino-terminal portion of Myb's carboxy-terminal domain. Leucine zippers, such as those found in the transcription factors *jun, fos,* and *myc* are thought to facilitate the protein-protein interactions which permit heterodimerization of DNA-binding proteins. As noted above, such dimerization is thought to play a key role in regulating the transcriptional activity of these factors. A Myb-dimerizing binding partner has yet to be identified but Myb-Myb homodimerization, which likely occurs through its leucine zipper, does lead to loss of DNA binding and transactivation ability. Accordingly, one could reasonably postulate that Myb-driven gene transactivation might be regulated by the binding of additional protein partners in the leucine zipper domain of the carboxy terminus of the protein. Alternatively, as was true for the transactivating function of the protein, repressor activity might also require the interaction of Myb with other proteins.

Targeting the c-*myb* Gene

My laboratory has been using AS ODN since 1988 to understand the role of c-*myb* and other proto-oncogenes in normal and malignant human hematopoiesis. These investigations were initially designed to elucidate the role of Myb protein in regulating hematopoietic cell development. Because the results obtained from these studies had obvious clinical relevance, more translationally oriented studies were undertaken. These have now culminated in Phase I clinical trials which are presently ongoing at the Hospital of the University of Pennsylvania. Below I summarize the development of the c-*myb*-targeted AS ODN and allude briefly to our initial clinical experience with this agent.

In Vitro Experience in the Hematopoietic Cell System

Role of c-*myb*-Encoded Protein in Normal Human Hematopoiesis. Attempts to exploit the c-*myb* gene as a therapeutic target for AS ODNs began as an outgrowth of studies that were seeking to define the role of Myb protein in regulating normal human hematopoiesis. During the course of these studies it was determined that exposing normal bone marrow mononuclear cells (MNC) to c-*myb* antisense oligomers resulted in a decrease in cloning efficiency and progenitor cell proliferation. The effect was lineage indifferent because c-*myb* antisense DNA inhibited granulocyte-macrophage colony-forming units (CFU-GM), CFU-E (erythroid), and CFU-Meg (megakaryocyte). In contrast, c-*myb* sense oligomers had no consistent effect on hematopoietic colony formation when compared with growth in control cultures ($p = 0.778$, $p = 0.796$, $p = 0.375$ for CFU-GM, CFU-E, and CFU-Meg, respectively). Inhibition of colony formation was also dose related and sequence specific. If progenitor cells were exposed to 0.3, 1.4, 7, and 14 μM c-*myb* ODNs the sense oligomers did not significantly affect colony formation whereas the antisense oligomers at the concentrations above lead to the formation of 21 ± 3, 17 ± 1, 10 ± 2, and 4 ± 2 megakaryocyte colonies, respectively. AS ODNs at concentrations of 1.4, 7, and 14 μM led to the formation of 775 ± 127, 515 ± 9, and 244 ± 25 myeloid colonies. The requirement for Myb during normal hematopoiesis has also been confirmed in gene-targeting experiments carried out using the technique of homologous recombination. In other investigations, it was also determined that hematopoietic progenitor cells appeared to require Myb protein during specific stages of development, particular when they were actively cycling.

***Myb* Protein Is Also Required for Leukemic Hematopoiesis.** Because the c-*myb* antisense oligomers inhibited normal cell growth, we were also interested in determining their effect on leukemic cell growth. Can one reasonably postulate that aberrant c-*myb* expression or Myb function might play a role in carcinogenesis? Amplification of c-*myb* in acute myelogenous leukemia (AML) and c-*myb* overexpression in 6q syndrome has been reported, which suggests that altered c-*myb* expression may indeed play a role in leukemogenesis. The exact nature of this role remains to be defined. One potential mechanism is through Myb's ability to regulate hematopoietic cell proliferation. Indeed, regulating progression through the G1/S gate of the cell cycle seems to be a key function of c-*myb*. The mechanism whereby c-*myb* regulates cell proliferation is uncertain but might relate to its role in regulating a number of important cell cycle genes including c-*myc*, proliferating cell nuclear antigen and cdc2. Myb may also play a role in regulating differentiation. Evidence to support this hypothesis includes the observations that as primitive hematopoietic cells mature c-*myb* expression declines and that constitutive expression of c-*myb* usually, but not always, inhibits the ability of hematopoietic cells to undergo differentiation. More directly, c-*myb* has been shown to function as a transcription factor for several cellular genes including the neutrophil granule protein mim-1, CD4, IGF-1, and CD34, and possibly other growth factors or growth factor receptors such as c-*kit*.

To address the question of Myb's role in leukemic hematopoiesis, we used a variety of leukemic cell lines, including those having a myeloid and lymphoid origin. In addition, we also used primary patient material. We first determined the effect of *myb* sense and antisense oligomers on the growth of HL-60, K562, KG-1, and KG-1a cell lines. The antisense oligomer inhibited the proliferation of each leukemia cell line, although the effect was most pronounced on HL-60 cells. Specificity of this inhibition was demonstrated by the fact that the sense oligomer had no effect on cell proliferation, nor did antisense sequences with two or four nucleotide mismatches. To determine whether the treatment with c-*myb* antisense modified cell cycle distribution of HL-60 cells, we measured the DNA content in exponentially growing cells exposed to either sense or antisense *myb* oligomers. Control cells and cells treated with c-*myb* sense oligomers had twice the DNA content of HL-60 cells exposed to the antisense oligomers. Most of these cells appeared to reside either in the G1 compartment or were blocked at the G1/S boundary. To examine the effect of the c-*myb* oligomers on lymphoid cell growth we used a lymphoid leukemia cell line, CCRF-CEM. As we noted in normal lymphocytes, the CCRF-CEM cells were extremely sensitive to the antiproliferative effects of the c-*myb* antisense DNA. When exposed to the sense oligomers, we found negligible effects on CEM cell growth in short-term suspension cultures. In contrast, exposure to c-*myb* antisense DNA resulted in a daily decline in cell numbers. Compared with untreated controls, antisense DNA inhibited growth ~2 logs. Growth reduction was not a cytostatic effect because cell viability was reduced only ~70% after exposure to the antisense oligomers, and CEM cell growth did not recover when cells were left in culture for an additional 9 days. The c-*myb* antisense oligomers also appeared effective in inhibiting cell growth of primary patient material when derived from patients with both acute and chronic myeloid leukemias. Colony inhibition was apparently quite efficient, approximately 58 to 95%, as determined by elimination of bcr-abl expressing chronic myelogenous leukemia (CML) clones.

Normal and Leukemic Progenitor Cells Rely Differentially on c-*myb* Function. To be useful as a therapeutic target, leukemic cells would have to be more dependent on Myb protein than their normal counterparts. To examine this critical issue we incubated phagocyte and T-cell-depleted normal human marrow MNC, human T-lymphocyte leukemic cell line blasts (CCRF-CEM), or 1:1 mixtures of these cells with sense or antisense ODN to codons 2–7 of human c-*myb* mRNA. Oligomers were added to liquid suspension cultures at time 0 and at time +18 hr. Control cultures were untreated. In controls, or in cultures to which "high" doses of sense oligomers were added, CCRF-CEM proliferated rapidly, whereas MNC numbers and viability decreased <10%. In contrast, when CCRF-CEM were incubated for 4 days in c-*myb* antisense DNA, cultures contained $4.7 \pm 0.8 \times 10^4$ cells/mL (mean ± SD; $n = 4$) compared with $285 \pm 17 \times 10^4$ cells/mL in controls. At the effective antisense dose, MNC were largely unaffected. After 4 days in culture, remaining cells were transferred to methylcellulose supplemented with recombinant hematopoietic growth factors. Myeloid colonies/clusters were enumerated at day 10 of culture inception. Depending on cell number plated, control MNC formed from 31 ± 4 to 274 ± 18 colonies. In dishes containing equivalent numbers of untreated or sense oligomer-exposed CCRF-CEM colonies were too numerous to count. When MNC were mixed 1:1 with CCRF-CEM in antisense oligomer concentrations ≤ 5 μg/mL, only leukemic colonies could be identified by morphologic, histochemical, and immunochemical analysis. However, when antisense oligomer exposure was intensified, normal myeloid colonies could now be found in the culture whereas leukemic colonies could no longer be identified with certainty using the same analytic methods. Finally, at antisense DNA doses used in the above-mentioned studies, AML blast cells from 18 of 23 patients exhibited ~75% decrease in colony and cluster formation compared with untreated or sense oligomer-treated controls. When 1:1 mixing experiments were carried out with primary AML blast cells and normal MNC, we were again able to preferentially eliminate AML blast colony formation while normal myeloid colonies continued to form.

Use of c-*myb* Oligomers as Bone Marrow Purging Agents. The experiments above suggested that leukemic cell growth could be preferentially inhibited after exposure to c-*myb* AS ODNs. It is compelling to contemplate a clinical use for our findings in bone marrow transplantation. In this application, exposure conditions are entirely under the control of the investigator. In addition, the patient's exposure to the antisense DNA is minimal. This circumstance also would make approval by regulatory agencies less difficult. We therefore determined whether the antisense oligomers could be used as *ex vivo* bone marrow-purging agents. To test this hypothesis, normal MNC were mixed (1:1) with primary AML or CML blast cells and then exposed to the oligomers using a slightly modified protocol designed to test the feasibility of a more intensive antisense exposure. With this in mind, an additional oligomer dose (20 μg/mL) was given just before plating the cells in methylcellulose. In control growth factor-stimulated cultures leukemic cells formed 25.5 ± 3.5 (mean ± SD) colonies and 157 ± 8.5 clusters (per 2×10^5 cells plated). Exposure to c-*myb* sense oligomers did not significantly alter these numbers (19.5 ± 0.7 colonies and 140.5 ± 7.8 clusters; $p > 0.1$). In contrast, equivalent concentrations of antisense oligomers totally inhibited colony and cluster formation by the leukemic blasts. Colony formation was also inhibited in the plates containing normal MNC, but only by ~50% in comparison with untreated control plates (control colony formation, 296 ± 40 per 2×10^5 cells plated; treated colony formation, 149 ± 15.5 per 2×10^5 cells).

Evaluating the Efficiency of an Antisense Purge. To assess the potential effectiveness of an antisense purge in CML patients, we carried out co-culture studies with cells obtained from patients in blast crisis and in the chronic phase of their disease. RNA was extracted from cells cloned in methylcellulose cultures after exposure to the highest c-*myb* antisense oligomer dose. The RNA was then reverse transcribed and the resulting cDNA amplified. For each patient studied, mRNA was also extracted from a comparable number of cells derived from untreated control colonies using the same technique. Eight cases were evaluated and in each case bcr-abl expression, as detected by reverse transcription-polymerase chain reaction (RT-PCR), correlated with colony growth in cell culture. In cases that were inhibited by exposure to c-*myb* antisense oligomers (7/11), bcr-abl expression was also greatly decreased or nondetectable. These results suggested that bcr-abl expressing CFU might be substantially or entirely eliminated from a population of blood or marrow MNC by exposure to the AS ODN. To explore this possibility further, replating experiments were conducted on samples from two patients. We hypothesized that if CFU belonging to the malignant clone were present at the end of the original 12-day culture period, but not detectable because of failure to express bcr-abl, they might re-express the message upon regrowth in fresh cultures. Accordingly, cells from these patients were exposed to oligomers and then plated into methylcellulose cultures formulated to favor growth of either CFU-GM or CFU-GEMM. As found with the original specimens, untreated control cells and cells exposed to sense oligomers had RT-PCR-detectable bcr-abl transcripts. Those exposed to the c-*myb* antisense oligomers had none. One of the paired dishes from these cultures was solubilized with fresh medium, and all cells contained therein were washed, disaggregated, and replated into fresh methylcellulose cultures without re-exposing the cells to oligomers. After 14 days, CFU-GM and CFU-GEMM colony cells were again probed for bcr-abl expression. Control and sense-treated cells had RT-PCR-detectable mRNA but none was found in the antisense-treated colonies. These results suggest that elimination of bcr-abl expressing cells and CFU was highly efficient and perhaps permanent.

Efficacy of c-*myb* Oligodeoxynucleotides *in Vivo*. The studies described above were carried out primarily with unmodified DNA. Such molecules are subject to endo- and exonuclease attack at the phosphodiester bonds and are therefore of little utility *in vivo*. Chemical modifications of ODN have sought to address this problem. For example, substitution of a sulfur for a hydrogen atom creates a class of compounds called phosphorothioates. These have several desirable properties including increased stability, water solubility, and ability to provide all RNase H substrate. To examine the *in vivo* effectiveness of phosphorothioate ODN, we established human leukemia/SCID (severe combined immunodeficiency) mouse chimeras and then treated the animals with the nuclease-resistant phosphorothioate oligomers targeted to the c-*myb* gene. SCID mice were injected intravenously with K562 chronic myeloid leukemia cells after cyclophosphamide conditioning. K562 cells express c-*myb*, the AS ODN target, and the tumor-specific bcr-abl oncogene which was used for tracking the human leukemia cells in the mouse host. After tumor cell injection, animals developed blasts in the peripheral blood within 4 to 6 weeks. After peripheral blood blast cells appeared, mean survival of untreated mice ($n = 20$) was 6 ± 3 (SD) days. Dying animals had prominent central nervous system infiltration, marked infiltration of the ovary, and scattered abdominal granulocytic sarcomas. Infusion of either sense or scrambled sequence c-*myb* phosphorothioate oligomers (24 base pairs; codons 2–9) for 3, 7, or 14 days had no statistically significant effect on sites of disease involvement or animal survival compared with control animals. In contrast, animals treated for 7 to 14 days with c-*myb* AS oligomers survived 3.5 to 8 times longer ($p < 0.001$) than the various control animals ($n = 60$). In addition, animals receiving c-*myb* AS DNA had either rare microscopic foci or no obviously detectable central nervous system disease, and a >50% reduction of ovarian involvement. A 3-day infusion of c-*myb* AS (100 μg/day) was without effect. Infusing mice ($n = 12$) with AS oligomers (200 μg/day × 14 days) complementary to the c-*kit* proto-oncogene, which K562 cells do not express, also had no effect on disease burden or survival ($n = 12$). These results suggested that phosphorothioate-modified c-*myb* antisense DNA might be efficacious for the treatment of human leukemia *in vivo*.

The c-*kit* Gene

Encouraged by our success with targeting the *myb* gene, we have attempted to apply the strategy outlined above to investigating the biology of other important hematopoietic genes and determining their suitability as potential therapeutics targets. Our studies on the c-*kit* receptor is a good example of this approach. The c-*kit* proto-oncogene is the normal cellular homologue of v-*kit*, the HZ4 feline sarcoma virus oncogene. The c-*kit* proto-oncogene encodes a dimeric transmembrane tyrosine kinase receptor homologous to the colony-stimulating factor (CSF)-1 and platelet-derived growth factor receptors. The mouse c-*kit* gene has been mapped to chromosome 5 where it was determined to be allelic with the dominant white spotting locus (W). W mice have a variety of abnormalities affecting coat color, gonadal development, and hematopoiesis. Hematopoiesis is characterized by a marked reduction in the number of erythroid burst-forming units (BFU-E) and a profound macrocytic anemia. The molecular lesions responsible for these defects are now being defined. For example, the W^{42} mutation is associated with particularly severe hematologic manifestation which is caused by a missense mutation leading to replacement of one amino acid and defective tyrosine kinase activity. The ligand for c-*kit* has also been identified recently. The encoded protein, known as stem cell factor, mast cell growth factor, or steel factor is the product of a gene that resides at the steel (Sl) locus. Mice with Sl mutations have phenotypic abnormalities quite similar to those of W mice, and the explanation for this curiosity is now apparent. The W mouse lacks, or has defects in a critical signal-transducing receptor encoded by c-*kit*, whereas the Sl mouse has defects in the ligand which stimulates the c-*kit* receptor. These observations establish the critical importance of the c-*kit* receptor in hematopoietic stem cell development.

Targeting the c-*kit* Gene in Normal Hematopoietic Cells. We first determined whether c-*kit* expression in MNC could be abrogated in a sequence-specific manner by exposure to c-*kit* AS ODN. MNC were incubated at 4°C for 24 hr in a low-serum-containing medium. Cells were rewarmed to 37°C and exposed to growth factors as described above in the presence of c-*kit* oligomers. Thirty-six hours after warming, c-*kit* mRNA was detected in control cells and in cells exposed to sense and scrambled sequence oligomers. In contrast, antisense-treated cells had no detectable c-*kit* mRNA. Accordingly, inhibition of c-*kit* expression was highly efficient and sequence specific.

The effect of c-*kit* antisense and control sequence oligomers on normal hematopoietic progenitor cell cloning efficiency and development was then investigated systematically by assessing the effect of oligomer exposure on CFU-E, BFU-E, CFU-GM, and CFU-MEG derived colony growth. The c-*kit* antisense oligomers inhibited erythroid colony formation in a dose-dependent fashion. Inhibition was sequence specific because neither sense nor scrambled sequence oligomers significantly affected colony growth compared with untreated controls. At the highest doses used, nonspecific inhibition of erythroid colony formation was sometimes seen, but this inhibition was never as great as that observed with the antisense oligomers. These inhibited growth of CFU-E and BFU-E, ~75% and ~71%, respectively. Residual colonies were much smaller than those of the untreated controls. Nevertheless, maturation of the cellular constituents of the colonies appeared normal in terms of the cells' ability to synthesize hemoglobin. Indirect immunofluorescence analysis of cell hemoglobin content using antibodies recognizing hemoglobins F and A revealed no discernible difference in the type or percentage of hemoglobin synthesized by the antisense-treated cells (data not shown). In contrast to the erythroid colony results, CFU-GM- and CFU-MEG-derived colony formation was unaffected by exposure to any of the oligomers, at any of the doses used.

To provide additional proof that c-*kit* antisense-mediated inhibition of erythropoiesis was caused by the absence of c-*kit* receptor function, we sought to demonstrate that BFU-E responsiveness to SCF could be abolished in a sequence-specific manner after exposure to c-*kit* oligomers. Accordingly CD34+ MNC (2×10^4) were cloned in the presence of erythropoietin (5 U/mL) and stem cell factor (100 ng/mL) alone, or with sense, antisense, or scrambled sequence c-*kit* oligomers (final concentration = 150 μg/mL = ~26 μM). In four experiments, 191 ± 19 BFU-E (mean ± SD) were grown in the presence of the growth factors alone. These numbers were not statistically different from those cloned with sense (183 ± 29; $p = 0.654$) or scrambled sequence oligomers (180 ± 20; $p = 0.758$). In the presence of the c-*kit* antisense oligomers BFU-E-derived colony formation was abolished completely (0.4 ± 0.7; $p < 0.0001$) suggesting that *kit* receptor was no longer present to interact with its ligand.

Targeting the c-*kit* Gene in Malignant Hematopoietic Cells. The data presented above suggested that the c-*kit*-encoded receptor plays an important role in normal hematopoiesis, especially for erythroid cell development. The c-*kit* mRNA and protein has been detected in some acute leukemias, which suggests a role for the *kit*-encoded protein in malignant hematopoietic cell growth as well. Nevertheless, its importance in the pathogenesis of these diseases was unclear because expression in primary leukemias is quite variable and, to our knowledge, has not been examined in other malignant hematopoietic diseases such as the myeloproliferative disorders. A^-T^- MNC were therefore obtained from patients with a variety of hematologic malignancies and exposed to the c-*kit* oligomers. The effect of oligomer exposure on the ability of malignant CFU-GM to form colonies in semisolid media was then assessed. Twenty patients were studied: three with acute lymphocytic leukemia, four with acute myelogenous leukemia, ten with chronic myelogenous leukemia, and four with polycythemia vera. CFU-GM from ten of these patients were inhibited by exposure to the c-*kit* antisense oligomers. Of these, eight were derived from patients with myeloproliferative disorders (all of whom had PCR-documented expression of c-*kit* mRNA; data not shown) suggesting that progenitor cells isolated from patients with these disorders were more dependent on c-*kit* function than those with acute leukemia. Five responding patients had CML. Interestingly, when residual colonies were probed for bcr-abl expressing cells by RT-PCR, bcr-abl expression was decreased substantially or undetectable in cases in which colony inhibition was observed. In patients with polycythemia vera, both BFU-E (92 ± 3%) and CFU-GM-derived colony formation were inhibited significantly by the c-*kit* antisense oligomers.

Antisense Oligonucleotides Move into the Clinic

Based on the type of data presented above, a favorable therapeutic index in toxicology testing, and more detailed knowledge of the pharmacokinetics of oligonucleotides, antisense oligonucleotides have begun to be evaluated in the clinic. These materials are presently being evaluated as topical agents in the treatment of herpes keratitis and genital warts and as systemic antiviral agents in the case of patients with human immunodeficiency virus. They are also being evaluated as anticancer agents. In this regard, clinical trials with AS ODN have begun at the University of Nebraska, the University "La Sapienza" in Rome, Italy, and at the University of Pennsylvania. All trials thus far have involved the treatment of hematologic malignancies, primarily acute or chronic myeloid leukemias (CML). We have initiated two such trials. One of these is a pilot study to evaluate the effectiveness of a phosphorothioate-modified ODN antisense to c-*myb* as a marrow-purging agent in patients with chronic phase (CP) or accelerated phase (AP) CML. The other is

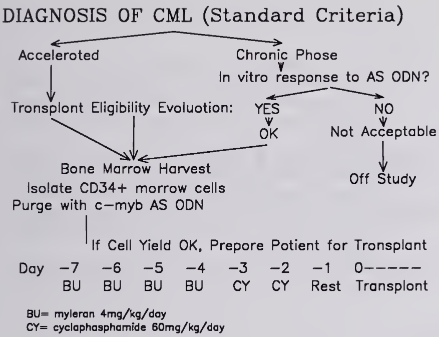

DIAGNOSIS OF CML (Standard Criteria)
Accelerated
Chronic Phase
In vitro response to AS ODN?
Transplant Eligibility Evaluation:
YES
OK
NO
Not Acceptable
Off Study
Bone Marrow Harvest
Isolate CD34+ marrow cells
Purge with c–myb AS ODN
If Cell Yield OK, Prepare Patient for Transplant
Day −7 −6 −5 −4 −3 −2 −1 0-----
BU BU BU BU CY CY Rest Transplant
BU= myleran 4mg/kg/day
CY= cyclophosphamide 60mg/kg/day

Fig. 1. Purging with c-*myb* AS ODN. Evaluation decision tree and treatment protocol outline.

a Phase I/II study evaluating the toxicity of systemically infused c-*myb* AS ODN in patients with CML in blast crisis or refractory acute leukemia.

The general outline of the purging protocol is shown in Figure 1. To purge the CML Marrows, $\geqslant 0.5 \times 10^6$ CD34+ MNC/kg are collected from the harvested marrow with a CD34 affinity column. Purging is carried out for 24 hr at 37°C with c-*myb* AS phosphorothioate ODN (AS PODN) followed by cryopreservation of the purged cells. Patients then receive myeloablative therapy with myleran (4 mg/kg/day for 4 days) and cyclophosphamide (60 mg/kg/day for 2 days), followed by reinfusion of P-ODN-purged MNC. G-CSF is administered concurrently. Eight patients have been treated thus far on this protocol and six appear to have derived hematologic and in some cases cytogenetic benefit from the procedure. The details of these cases are being prepared for publication.

Seventeen patients have been treated with the systemic infusions in a dose escalation study. No untoward toxicity has been encountered. The details of these cases are also being prepared for publication. In essence, however, although the follow-up period is very short, and total patient numbers are small, our results, and those preliminary results reported by our colleagues, suggest that purging with phosphorothioate-modified c-*myb* AS ODN is a safe and potentially useful procedure in patients who have leukemia without an allogeneic marrow donor.

Antisense DNA as Investigative Tool and Therapeutic Strategy—A Caveat

Although seemingly simple in theory and execution, antisense experiments may be far more difficult to conduct than anticipated; at times, reported results have been difficult to reproduce. The reasons why these difficulties arise may be placed into three major categories. First, problems with reproducibility can be explained by many technical factors including differences in oligomer synthesis, handling, storage, and purification. Differences in experimental conditions and sensitivity between subtypes of cell lines may also be factors. An additional important variable, as with any technique, is the experience of the research team using this strategy. Second, problems relating to the physicality of ODN target mRNA interactions are also difficult to predict. Cellular mRNAs are thought have highly ordered secondary and tertiary structures. If the targeted sequence is buried within the core of a structurally complicated mRNA molecule, or rendered inaccessible by an associated protein, then hybridization of the ODN and mRNA may not be possible. Generally, it is difficult to predict the secondary structure of mRNA molecules, and accessible regions must be determined empirically by synthesizing ODN to multiple sites and selecting those which are most effective. Another important problem for which the antisense strategy is criticized focuses on effects created by nonspecific interactions of the ODN with other mRNA targets or intracellular proteins. These problems also appear to be related to the chemistry, and concentration, of the oligo used. For example, there are some data that phosphorothioated AS ODN designed to inhibit viral proliferation may do so through nonsequence-dependent inhibition of viral polymerases or by nonspecific absorption onto cell membranes. These actions are thought to be the result of interactions between the phosphorothioated backbone and cell surface or viral proteins. Nevertheless, so-called nonspecific effects can at times be advantageous if appropriately exploited. The caveat here is not to confuse a biologic effect with an "antisense effect."

Suggested Readings

1. Agrawal S, Tang JY. GEM91: an antisense oligodeoxynucleotide phosphorothioate as a specific therapeutic agent for AIDS. Antisense Res Dev 1992; 2:261–266.
2. Bayever E, Iversen P, Smith L, et al. Guest editorial: systemic human antisense therapy begins. Antisense Res Dev 1992; 2:109–110.
3. Cohen JS, Hogan ME. The new genetic medicines. Sci Am 1994; 271:76–82.
4. Gewirtz AM. Therapeutic applications of antisense DNA in the treatment of human leukemia. Ann NY Acad Sci 1992; 660:178–187.
5. Gewirtz AM, Calabretta B. A c-myb antisense oligodeoxynucleotide inhibits normal human hematopoiesis in vitro. Science 1988; 245:1303–1306.

6. Hijiya N, Zhang J, Ratajczak MZ, et al. The biologic and therapeutic significance of c-Myb expression in human melanoma. Proc Natl Acad Sci USA 1993; 91:4499–4503.
7. Kitajima I, Shinohara T, Bilakovics J, et al. Ablation of transplanted HTLV-I Tax-transformed tumors in mice by antisense inhibition of NF-kb. Science 1992; 258: 1792–1795.
8. Milligan JF, Matteucci MD, Martin JC. Current concepts in antisense drug design. J Med Chem 1993; 36:1923–1937.
9. Ratajczak MZ, Kant JA, Luger SM, et al. In vivo treatment of human leukemia in a scid mouse model with c-myb antisense oligodeoxynucleotides. Proc Natl Acad Sci USA 1993; 89:11823–11827.
10. Simons M, Edelman ER, DeKeyser JL, et al. Antisense c-myb oligonucleotides inhibit intimal arterial smooth muscle cell accumulation in vivo. Nature 1992; 359:67–70.
11. Skorski T, Nieborowska-Skorska M, Barietta C, et al. Highly efficient elimination of Philadelphia leukemic cells by exposure to bcr/abl antisense oligodeoxynucleotides combined with mafosfamide. J Clin Invest 1993; 92:194–202.
12. Snyder DS, Wu Y, Wang JL, et al. Ribozyme mediated inhibition of bcr-abl gene expression in a Philadelphia chromosome-positive cell line. Blood 1993; 82:600–605.
13. Stein CA, Cheng YC. Antisense oligodeoxynucleotides as therapeutic agents: is the bullet really magical? Science 1993; 261:1004–1012.
14. Wagner, RF. Gene inhibition using antisense oligodeoxynucleotides. Nature 1994; 372:333–335.

G Proteins: Out of the Cytoskeletal Closet

Martin Rodbell, Ph.D.

Abstract

This article contains a brief review of GTP-binding proteins (G protein) signaling mechanism with emphasis on accumulated information which suggests that G proteins are multimeric proteins structured such that one receptor can catalytically activate each of the monomers as it moves in an oscillatory fashion along the multimeric chain. Movement is dictated by the binding of GTP or GDP controlled by the exchange reaction induced by agonist binding to the receptor. Based on the dynamic instability model for the interactions of myosin and F-actin, the hypothesis is presented that a GTP-bound monomer is released from one end of the multimer allowing it to interact with effectors such as adenylyl cyclase embedded in the plasma membrane. Association with the enzyme results in a transition, state of the enzyme-G protein complex. In the presence of magnesium, the GTPase on the α-subunit of Gs (which stimulates adenylyl cyclase) is activated, resulting in the interaction of the α- and βγ-subunits of Gs with different domains of adenylyl cyclase. In this fashion, the GTPase is not simply a turnoff mechanism, but induces a new configuration of the G protein-cyclase complex allowing for greatly enhanced cyclic AMP formation. Dissociation of bound Pi may be one of the rate-limiting steps allowing for cycling of G proteins between the multimer-receptor complex and adenylyl cyclase.

GTP-BINDING PROTEINS (G proteins) play a crucial role in transmembrane signaling (1, 2). Characterized as heterotrimeric structures, G proteins consist of at least 21 different GTP-binding units (α-subunits), 5 β-subunits, and 6 γ-subunits (3). They couple with more than 300 different types of receptors, including the complex array of olfactory receptors (4). Typified by having seven membrane-spanning domains, functional coupling of these receptors requires all three subunits of G proteins; acylation of the α- and γ-subunits with fatty acyl and isoprenyl groups plays a role in membrane attachment and the coupling process (5, 6).

Still unresolved are the topological organization of G proteins, receptors, and the affected signaling devices (phosphodiesterases, phospholipases, adenylate cyclase, ion channels, etc.) in the surface membrane and how these components are linked during the process of signal transduction. Two prominent models (2, 7) invoke the concept of receptor mobility as a basis for explaining the accumulated evidence that a single receptor can activate several G proteins. In these theories, binding of agonists to the receptors promotes coupling with G proteins. Although freely mobile receptors can be detected, there is abundant evidence for precoupled receptors (8–13), for interactions of receptor-coupled G proteins with the cytoskeletal network (14–22), and for clustered aggregates of receptors and G proteins (23). Several G protein-regulated effectors, including adenylyl cyclase and ion channels, appear to be associated with

---

Adapted from a Dean's Lecture presented at The Mount Sinai School of Medicine on October 11, 1995.

From the Signal Transduction Section, Laboratory Molecular and Cellular Pharmacology National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina. Address reprint requests to Martin Rodbell, Ph.D., Bldg. 18-01, NIEHS, Research Triangle Park, NC 27709.

actin, actin-binding proteins, or other types of cytoskeletal proteins (24–29). These findings imply organized, fixed structures somehow associated with the intracellular cytoskeletal matrix. Finally, there is indirect evidence (30) for compartmentation in the same cell of different receptor types coupled to the same Gs, the stimulator of adenylyl cyclase.

Faced with the probability that receptor-coupled G protein systems and their effectors are not mobile structures, how can each receptor-regulated system be compartmentalized, essentially immobile within the plasma membrane, but the signals are produced in rapidly (fractions of seconds) reversible fashion? What process or structure brings about such rapid, expeditious communication between components?

The "disaggregation" theory of signal transduction (31), with recent modifications (1), supplies some potential answers. In the original concept, based on the powerful technique of target irradiation analysis (32), receptors assemble with G proteins as large multimeric structures ($>10^6$ daltons) which, when activated by hormones, GTP, and Mg^{2+}, release "monomers" of proteins; the monomers, depending on the type of receptor and G protein, cause either stimulation or inhibition of effectors such as adenylyl cyclase. From its inception, the model incorporated the notion that receptors and its G protein couple stoichiometrically. However, accumulative evidence indicates that, with few exceptions, heterotrimeric G proteins exist far in excess (10–100 times) of receptors in the cell membrane. Accordingly, it can be deduced that the functional mass given by irradiation analysis comprises multimeric structures of G proteins. With this assumption, for example, the functional mass of Gs coupled to the glucagon receptor, as determined by irradiation analysis, is 800 kDa (33). Given the known mass of Gs (90–95 kDa), a multimer of Gs coupled to the glucagon receptor consists of 8 to 10 monomer (heterotrimers) units; i.e., one receptor activates a number of monomers in the attached multimer, consistent with the catalytic actions of receptors. Supportive biochemical evidence for multimeric structures stems from a series of studies using octylglucoside or digitonin to extract G proteins from brain and liver membranes (34–36). G proteins extracted with these detergents display hydrodynamic properties suggestive of structures that are considerably larger than observed in cholate and lubrol, the detergents normally used for extraction and purification of heterotrimeric G proteins. Important structural evidence that G proteins are multimeric in their attachment to biological membranes is that all of the endogenous G protein types in rat brain synaptoneurosomes (Gs, Gi, Go, Gq) can be cross-linked to yield structures similar in size to cross-linked, multimeric tubulin associated with the same membranes (37). G proteins also form very large structures after extraction with digitonin (36).

The "disaggregation" theory predicts that hormones or agents that bypass the receptor-activation process (metal ion complexes with fluoride ion, nonhydrolyzable analogs of GTP) stimulate the release of monomers (heterotrimers) as suggested from the functional target mass (subtracting the contribution of the mass given by adenylyl cyclase (38)) obtained after preactivation of the enzyme (33). Biochemical evidence obtained with preactivated adenylyl cyclase in brain and turkey erythrocyte membranes also indicates that heterotrimeric Gs (with bound GTPγS) is attached to the enzyme (7). In addition, glucagon, combined with GTPγS and Mg^{2+}, activates a large structure containing Gs, yielding monomeric Gs (35). Similarly, activation of several species of G proteins in brain synaptoneurosome by GTPγS and Mg^{2+} also produces monomers from larger structures, the relative effectiveness of the nucleotide being $Gi = Go \gg Gs > Gq$ (36).

A hypothetical reaction scheme (adapted from Reference 1) depicting interactions between receptor and multimeric Gs is shown in Figure 1A. The fundamental premise is that a receptor couples to a multimeric chain of G proteins and moves along the chain in an oscillatory fashion. The frequency and strength of attachment is determined by tightly linked hormone-occupied receptors typified by the low dissociation rates of agonists and by loosely coupled receptors (low affinity for agonists) when associated G proteins are occupied by guanine nucleotides (i.e., receptors coupled to nucleotide-free G proteins form strong attachments, receptors coupled to G proteins containing bound nucleotides form weak attachments). These reciprocal relationships in coupling (39–41) provide the means for receptors to move in an oscillatory fashion along the multimeric chain in a fashion similar to the nucleotide-dependent movement of myosin along the chain of F-actin. In addition, it should be noted that GDP is equal to or more effective than GTP in reducing receptor affinity (42). Hence, during its excursion along the chain, the receptor, accessible to at least two binding states for agonist and three states with coupled G proteins, intermittently forms strong and weak states of coupling. Potentially, depending on the relative concentrations of

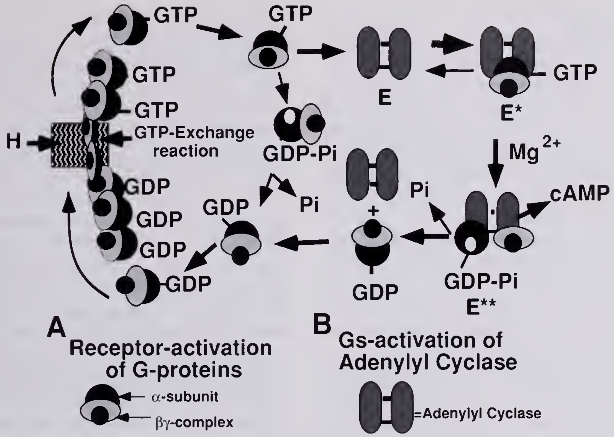


Fig. 1. Model for receptor activation of G proteins. All components remain bound to the surface membrane during the entire process. G proteins contained in the multimer are depicted in three states: some with bound GDP, some unoccupied by nucleotide (tightly interacted with high-affinity, hormone (H)-occupied receptor), and some occupied by GTP. Coupling of receptor to the chain is stochastic with relative coupling efficiency of each monomer being a function of the amount and type of nucleotide bound: unbound ≫ GTP > GDP. Loose and tight coupling of the receptor during its transit along the chain is the result of the differing affinity states of the G proteins (monomers) in the chain. The nucleotide uptake or exchange reaction occurs at each monomer unit irrespective of loose or tight coupling. As depicted, the most efficacious activation occurs with the high-affinity, tightly coupled form of the receptor. Release of monomers occupied by GTP from multimers is depicted in "quantal" fashion from one end of the multimer. In this scheme hydrolysis of GTP by GTPase and release of Pi are factors that control cycling between monomers and multimers. See text for other details.

GTP and GDP within the vicinity of the transduction process (see below), each of the monomers can be occupied and activated by GTP. In this theory, release of one or more GTP-bound monomers is fundamental to the activation of effectors such as adenylyl cyclase, ion channels, and phospholipases or phosphodiesterases.

An interesting outgrowth of this concept is the possibility of explaining the growing evidence that a single receptor not only catalytically activates several of the same species of G proteins but also more than one species of G proteins. The latter is exemplified with transfected receptors for luteinizing hormone and parathyroid hormone. Each receptor induces both increased cyclic AMP production and increased release of calcium ions from internal stores through the inositol-triphosphate pathway (43–45). This phenomenon is explained if the multimers contain two or more types of G proteins, each type regulating different types of effectors. To date, no other reaction scheme provides an explanation for this phenomenon.

Hydrolysis of GTP Is Fundamental to Physiological Actions of Heterotrimeric G Proteins

A dominant theme in signaling mechanisms is that binding of GTP induces action, whereas hydrolysis of GTP leads to inaction. This "turnon, turnoff" mechanism, however, is an oversimplification of what appears to be a multistep process. For example, modeling of the kinetics of activation of the adenylyl cyclase system in liver membranes by GTP and Mg^{2+} led to the transition-state concept (46) in which hydrolysis of GTP to GDP + Pi is intimately associated with activation of the enzyme. The model, schematically depicted in contemporary structural form (Fig. 1B), is defined by the basal state (E) which is converted reversibly to the transition state (E*) when interacting with GTP-bound Gs. Two magnesium-dependent processes drive conversion to the high-activity state (E**): (i) hydrolysis of GTP to GDP and Pi and (ii) dissociation of Gs into α_s

and $\beta\gamma$ (47). Release of Pi accelerates return of the system to E. α_s and $\beta\gamma$ are depicted as binding to separate domains or "cassettes" on adenylate cyclase (38). In this manner, α_s and $\beta\gamma$ activate independently but synergistically certain species of adenylyl cyclase (48), whereas $\beta\gamma$ inhibits other types (49). Other effectors may behave similarly. For example, miniglucagon regulates a calcium pump in a biphasic manner in hepatic membranes using synergistically $\beta\gamma$ and a cholera toxin-activated α_s-protein (50). Hence, the possible confirmations available to an effector may dictate whether the α- and $\beta\gamma$-subunits exert positive or negative regulatory effects on an effector; in either case the intact heterotrimeric protein initiates the signaling process.

As described recently (51), on-site hydrolysis of GTP to GDP and Pi may be responsible for initiating the rapid onset of activation of cyclic GMP phosphodiesterase, a major recipient of the photoactivation process in rod outer segments. The γ-subunit of the recipient cyclic GMP phosphodiesterase promotes rapid release of Pi, a rate-limiting "turnoff" step (52). Thus, although derived from different perspectives, both the activation of adenylyl cyclase by Gs and the activation of phosphodiesterase by Gt, or transducin, appear identical in their overall dynamic behavior. The recent finding that phospholipase C_β also induces rapid turnover of GTP when interacting with Gq (or G_{11}) indicates this may be a general phenomenon (53, 54). Hence, as measured generally, Pi release by isolated G proteins is not necessarily a measure of the rate of on-site GTP cleavage and appears to be a slow process relative to the activating hydrolytic cleavage (51). The long-lasting effects of fluoride ion (complexed with magnesium or aluminum ions), in contrast to GTP, are caused by its much slower egress from the Pi binding domain. As for the nonhydrolyzable analogs of GTP (Gpp (NH) p or GTPγS), their slow onset of activation relative to that of GTP (essentially instantaneous) and their pseudoirreversible effects may reflect hysteretic stabilization of the activated configuration produced by Mg^{2+}-induced dissociation of G proteins.

Conclusions

The concept of signal transduction via G proteins presented here raises aspects of cellular regulation hitherto confined to cytoskeletal proteins. As discussed elsewhere in detail (1), the network of cytoskeletal proteins is linked to surface membrane proteins and relays "information" from the external to the internal aspects of the cell through its structural changes. In this respect, multimeric G proteins linked to membrane receptors share similar structural and regulatory properties and can be considered components of the cytoskeletal network, as inferred from previous studies (55, 56). To account for the rapid actions of hormones on several protein species located within fixed, spatially distinct domains in a membrane, monomeric G proteins are the likely major communicating elements. Attached to phospholipids along the innerface of the surface membrane through lipophilic, isoprenylated γ-subunits (6), these proteins are ideal mobile vehicles for rapid communication. Dissociation and release of α-subunits to the cytosol (57, 58), densely packed with protein (59) and hence too viscous for rapid diffusion, seems unlikely as a means of rapid communication and may reflect some unknown aspect of the signaling process. In this respect, a major defect in the theory is the lack of information on the control of both the cycling and interactions between monomers (and their dissociated products) and multimers. A determining factor may be the relative amounts of GTP and GDP available at the critical sites of transduction. It is interesting that ATP/ADP gradients seem to be important for assembly/disassembly of actin filaments (60). Nucleoside diphosphate kinase is capable, along with nucleotidases, of controlling the relative concentrations of GTP and GDP. Perhaps the controversial role of the phosphokinase in G protein regulation (61) may be resolved if this enzyme, possibly concentrated at the sites of transduction (62), contributes to G protein regulation (61, 62) by controlling a gradient of GTP and GDP within the vicinity of the signaling process.

References

1. Rodbell M. The role of GTP-binding proteins in signal transduction: from the sublimely simple to the conceptually complex. Curr Top Cell Regul 1992; 32:1–47.
2. Gilman AG. G proteins: transducers of receptor-generated signals. Annu Rev Biochem 1987; 56:615–649.
3. Simon MI, Strathmann MP, Gautam N. Diversity of G proteins in signal transduction. Science 1991; 252:802–808.
4. Levy NS, Bakalyar HA, Reed RR. Signal transduction in olfactory neurons. J Steroid Biochem Mol Biol 1991; 39:633–638.
5. Linder ME, Middleton P, Hepler JR, et al. Lipid modifications of G proteins: alpha subunits are palmitoylated. Proc Natl Acad Sci USA 1993; 90:3675–3679.
6. Sanford J, Codina J, Birnbaumer L. Gamma-subunits of G proteins, but not their alpha- or beta-subunits, are polyisoprenylated. Studies on posttranslational modifications using in vitro translation with rabbit reticulocyte lysates. J Biol Chem 1991; 266:9570–9579.
7. Bar-Sinai A, Marbach I, Shorr RG, Levitzki A. The Gpp-

NHp-activated adenylyl cyclase complex from turkey erythrocyte membranes can be isolated with its beta gamma subunits. Eur J Biochem 1992; 207:703–708.

8. Jouneaux C, Audigier Y, Goldsmith P, et al. Gs mediates hormonal inhibition of the calcium pump in liver plasma membranes. J Biol Chem 1993; 268:2368–2372.
9. Nerme V, Severne Y, Abrahamsson T, Vauquelin G. Spontaneous coupling of the beta-adrenergic receptor to Ns in mammalian cardiac membranes. Mol Pharmacol 1986; 30:1–5.
10. Jagadeesh G, Tian WN, Deth RC. Agonist-induced modulation of agonist binding to alpha-1-adrenoceptors in bovine aorta. Eur J Pharmacol Mol Pharmacol 1991; (Sect 6):163–170.
11. Leung E, Jacobson KA, Green RD. Apparent heterogeneity of cardiac A1 adenosine receptors as revealed by radioligand binding experiments on N-ethylmaleimide treated membranes. Naunyn-Schmiedebergs Arch Pharmakol 1991; 344:639–644.
12. Neubig RR, Gantzos RD, Thomsen WJ. Mechanism of agonist and antagonist binding to alpha-2 adrenergic receptors: evidence for a precoupled receptor-guanine nucleotide protein complex. Biochemistry 1988; 27:2374–2384.
13. Watanabe T, Umegaki K, Smith WL. Association of a solubilized reactive prostaglandin E_2 receptor from renal medulla with a pertussis toxin-guanine nucleotide regulatory protein. J Biol Chem 1986; 261:13430–13439.
14. Kwatra MM, Schwinn DA, Schreurs J, et al. The substance-P receptor which couples to Gq/11 is a substrate of beta-adrenergic receptor kinase-1 and kinase-2. J Biol Chem 1993; 268:9161–9164.
15. Cantiello HF, Prat AG, Bonventre JV, et al. Actin binding protein contributes to cell volume regulatory ion channel activation in melanoma cells. J Biol Chem 1993; 268:4596–4599.
16. Carlson KE, Woolkalis MJ, Newhouse MG, Manning DR. Fractionation of the beta subunit common to guanine nucleotide-binding regulatory proteins with the cytoskeleton. Mol Pharmacol 1986; 30:463–468.
17. Wang N, Rasenick MM. Tubulin G protein interactions involve microtubule polymerization domains. Biochemistry 1991; 30:10957–10965.
18. Sarndahl E, Lindroth M, Bengtsson T, et al. Association of ligand-receptor complexes with actin filaments in human neutrophils: a possible regulatory role for a G-protein. J Cell Biol 1989; 109:2791–2799.
19. Roychowdhury S, Wang N, Rasenick MM. G-protein binding and G-protein activation by nucleotide transfer involve distinct domains on tubulin-regulation of signal transduction by cytoskeletal elements. Biochemistry 1993; 32:4955–4961.
20. Sarndahl E, Bokoch GM, Stendahl O, Andersson T. Stimulus-induced dissociation of alpha-subunits of heterotrimeric GTP-binding proteins from the cytoskeleton of human neutrophils. Proc Natl Acad Sci USA 1993; 90: 6552–6556.
21. Jesaitis AJ, Bokoch GM, Tolley JO, Allen RA. Lateral segregation of neutrophil chemotactic receptors into actin- and fodrin-rich plasma membrane microdomains depleted in guanyl nucleotide regulatory proteins. J Cell Biol 1988; 107:921–928.
22. Vaziri C, Downes CP. Association of a receptor and G protein regulated phospholipase C with the cytoskeleton. J Biol Chem 1992; 267:22973–22981.
23. Wang HY, Berrios M, Malbon CC. Localization of beta-adrenergic receptors in A431 cells in situ. Effect of chronic exposure to agonist. Biochem J 1989; 263:533–538.
24. Watson PA. Direct stimulation of adenylate cyclase by mechanical forces in S49 mouse lymphoma cells during hyposmotic swelling. J Biol Chem 1990; 269:6569–6975.
25. Levine HI, Sayhoun NE, Cuatrecasas P. Properties of rat erythrocyte membrane cytoskeletal structures produced by digitonin extraction: digitonin insoluble beta-adrenergic receptor, adenylate cyclase, and cholera toxin substrate. J Membr Biol 1982; 64:225–231.
26. Davis LH, Bennett V. Mapping the binding site of human erythrocyte ankyrin for the anion exchanger and spectrin. J Biol Chem 1990; 265:10589–10596.
27. Edelstein NG, Catterall WA, Moon RT. Identification of a 33-kilodalton cytoskeletal protein with high affinity for the sodium channel. Biochemistry 1988; 27:1818–1822.
28. Cantiello HF, Stow JL, Prat AG, Ausiello DA. Actin filaments regulate epithelial Na+ channel activity. Am J Physiol 1991; 261:C882–C888.
29. Rodbell M. The role of hormone receptors and GTP regulatory proteins in membrane transduction. Nature 1980; 284:17–22.
30. Kempner ES. Novel predictions from radiation target analysis. Trends Biochem Sci 1993; 18:236–239.
31. Schlegel W, Kempner ES, Rodbell M. Activation of adenylate cyclase in hepatic membranes involves interactions of the catalytic unit with multimeric complexes of regulatory proteins. J Biol Chem 1979; 254:5168–5176.
32. Nakamura S, Rodbell M. Octyl glucoside extracts GTP binding regulatory proteins from rat brain "synaptoneurosomes" as large, polydisperse structures devoid of beta gamma complexes and sensitive to disaggregation by guanine nucleotides. Proc Natl Acad Sci USA 1990; 87:6413–6417.
33. Nakamura S, Rodbell M. Glucagon induces disaggregation of polymer-like structures of the alpha subunit of the stimulatory G protein in liver membranes. Proc Natl Acad Sci USA 1991; 88:7150–7154.
34. Jahangeer S, Rodbell M. The disaggregation theory of signal transduction revisited—further evidence that G proteins are multimeric and disaggregate to monomers when activated. Proc Natl Acad Sci USA 1993; 90: 8782–8786.
35. Coulter S, Rodbell M. Heterotrimeric G proteins in synaptoneurosome membranes are crosslinked by *p*-phenylenedimaleimide yielding structures comparable in size to crosslinked tubulin and F-actin. Proc Natl Acad Sci USA 1992; 89:5842–5846.
36. Krupinski J, Coussen F, Bakalyar HA, et al. Adenylyl cyclase amino acid sequence: possible channel or transporter like structure. Science 1989; 244:1558–1564.
37. Rodbell M, Krans HM, Pohl SL, Birnbaumer L. The glucagon-sensitive adenyl cyclase system in plasma membranes of rat liver. IV. Effects of guanyl-nucleotides on binding of ^{125}I-glucagon. J Biol Chem 1971; 246:1872–1876.
38. Welton AF, Lad PM, Newby AC, et al. Solubilization and separation of the glucagon receptor and adenylate cyclase in guanine nucleotide-sensitive states. J Biol Chem 1977; 252:5947–5950.
39. Limbird LE, Gill DM, Lefkowitz RJ. Agonist-promoted coupling of the beta-adrenergic receptor with the guanine nucleotide regulatory protein of the adenylate cyclase system. Proc Natl Acad Sci USA 1980; 77:775–779.
40. Rojas FJ, Birnbaumer L. Regulation of glucagon receptor

binding: lack of effect of Mg and preferential role for GDP. J Biol Chem 1984; 260:7829–7835.

41. Eason MG, Kurose H, Holt BD, et al. Simultaneous coupling of alpha2-adrenergic receptors to two G proteins with opposing effects. Subtype-selective coupling of alpha2C10, alpha2C4, and alpha2C2 adrenergic receptors to Gi and Gs. J Biol Chem 1992; 267:15795–15801.
42. Abou-Samra AB, Juppner H, Force T, et al. Expression cloning of a rat bone PTH/PTHrP receptor: a single receptor activates both adenylate cyclase and phospholipase C. Proc Natl Acad Sci USA 1992; 89:2732–2736.
43. Gudermann T, Birnbaumer M, Birnbaumer L. Evidence for dual coupling of the murine luteinizing hormone receptor to adenylyl cyclase and phosphoinositide breakdown and Ca^{2+} mobilization. J Biol Chem 1992; 267:4479–4489.
44. Rendell MS, Rodbell M, Berman M. Activation of hepatic adenylate cyclase by guanyl nucleotides. Modeling of the transient kinetics suggests and "excited" state of GTPase is a control component of the system. J Biol Chem 1977; 252:7909–7912.
45. Codina J, Hildebrandt JD, Birnbaumer L, Sekura RD. Effects of guanine nucleotides and Mg on human erythrocyte Ni and Ns, the regulatory components of adenylyl cyclase. J Biol Chem 1984; 259:11408–11418.
46. Tang WJ, Gilman AG. Type-specific regulation of adenylyl cyclase by G protein beta-gamma subunits. Science 1991; 254:1500–1534.
47. Taussig R, Quarmby LM, Gilman AG. Regulation of purified type I and type II adenylyl cyclases by G protein beta-gamma subunits. J Biol Chem 1993; 268:9–12.
48. Lotersztajn S, Pavoine C, Deterre P, et al. Role of G protein beta-gamma subunits in the regulation of the plasma membrane calcium pump. J Biol Chem 1992; 267:2375–2379.
49. Ting TD, Ho YK. Molecular mechanism of GTP hydrolysis by bovine transducin: pre-steady-state kinetic analyses. Biochemistry 1991; 30:8996–9007.
50. Arshavsky VY, Bownds MD. Regulation of deactivation of photoreceptor G protein by its target enzyme and cGMP. Nature 1992; 357:416–417.
51. Maurice DH, Waldo GL, Morris AJ, et al. Identification of G-alpha-11 as the phospholipase C activating G-protein of turkey erythrocytes. Biochem J 1993; 290:765–770.
52. Berstein G, Blank JL, Smrcka AV, et al. Reconstitution of agonist-stimulated phosphatidylinositol 4,5-bisphosphate hydrolysis using purified M_1 muscarinic receptor, Gq/11, and phospholipase C $beta_1$. J Biol Chem 1992; 267:8081–8088.
53. Bourguignon L, Walker G, Huang HS. Interactions between a lymphoma membrane-associated guanosine 5′-triphosphate-binding protein and the cytoskeleton during receptor patching and capping. J Immunol 1990; 144:2242–2252.
54. Lokeshwar VB, Bourguignon L. The lymphoma transmembrane glycoprotein gp85 cd44 is a novel guanine nucleotide-binding protein which regulates gp85 cd44 ankyrin interaction. J Biol Chem 1992; 267:22073–22078.
55. Ransnas LA, Leiber D, Insel PA. Inhibition of subunit dissociation and release of the stimulatory G protein, Gs, by beta gamma-subunits and somatostatin in S49 lymphoma cell membranes. Biochem J 1991; 280:303–308.
56. Levis MJ, Bourne HR. Activation of the alpha subunit of G_s in intact cells alters its abundance rate of degradation and membrane avidity. J Cell Biol 1992; 119:1297–1307.
57. Minton AP, Colclasure GC, Parker JC. Model for the role of macromolecular crowding in regulation of cellular volume. Proc Natl Acad Sci USA 1992; 89:10504–10506.
58. Carlier MF, Jean C, Rieger KJ, et al. Modulation of the interaction between G-actin and thymosin beta 4 by the ATP/ADP ratio: possible implication in the regulation of actin dynamics. Proc Natl Acad Sci USA 1993; 90: 5034–5038.
59. Randazzo PA, Kahn RA, Northup JK. Nucleoside diphosphate kinase: conclusions withdrawn. Science 1992; 257:862–862.
60. Kimura N, Shimada N. Evidence for complex formation between GTP binding protein (Gs) and membrane-associated nucleoside diphosphate kinase. Biochem Biophys Res Commun 1990; 168:99–106.
61. Bominaar AA, Molijn AC, Pestel M, et al. Activation of G-proteins by receptor-stimulated nucleoside diphosphate kinase in dictyostelium. EMBO J 1993; 12:2275–2279.
62. Wieland T, Bremerich J, Gierschik P, Jakobs KH. Contribution of nucleoside diphosphokinase to guanine nucleotide regulation of agonist binding to formyl peptide receptors. Eur J Pharmacol 1991; 208:17–23.

Peptic Ulcer Disease: Paradigms Lost

Andrea Sarner, M.A. and Mark W. Babyatsky, M.D.[1]

Introduction

Although the incidence of peptic ulcer disease has steadily declined since the turn of the century in the United States, about 350,000 new cases are still diagnosed per year, leading to substantial health care costs, morbidity, and mortality (1). The conceptual focus for the etiology of peptic ulcer disease has evolved dramatically. In the early 20th century, psychosocial stress and dietary factors were considered the main etiologic agents; treatment relied on bed rest and a bland diet (2). In 1910, Schwarz coined the paradigm, "No acid, no ulcer" (3); acid remained the central paradigm of peptic ulcer disease throughout most of the century. In 1971, Sir James Black identified the histamine-2 (H2) receptor and demonstrated its importance in mediating gastric acid secretion (4). H2 receptor antagonists soon became a mainstay of ulcer therapy, leading to rapid healing of acute ulcers and lending credence to the Schwarz paradigm. However, ulcer recurrences remained frequent (5) and complication rates of bleeding and perforation were largely unaffected by acid blockade. Further, gastric acid secretory responses, the subject of intense investigation for many years, proved to overlap significantly in ulcer and normal subjects (6). Although acid remained the central paradigm through the 1970s, the physiology of pepsin was shown to play as indispensable a role as acid in the pathogenesis of peptic ulcer disease (7, 8).

Cytoprotection—Definition

As histamine receptor blockers were being introduced into ulcer treatment, a new concept in gastric mucosal defense was being formed. At the Upjohn Laboratories in Kalamazoo, Michigan, Andre Robert demonstrated prostaglandin protection of rat gastric mucosa against injury by a variety of experimental agents including acid, ethanol, and indomethacin (9), a property he termed "cytoprotection." Although prostaglandins inhibit gastric acid secretion through a cyclic AMP mechanism (10), Robert clearly showed that protection occurred at doses of prostaglandins significantly lower than those required to inhibit acid secretion (11). Further, acid blockade by cimetidine did not protect against these injuries. Because many agents other than prostaglandins have since been shown to protect the gastric mucosa, cytoprotection can now be defined as the ability of *pharmacological* agents to prevent or reduce gastric mucosal injury by mechanisms other than inhibition of acid secretion. Further, as other organs including the liver, pancreas, and distal gastrointestinal tract can also be protected against damage by agents such as prostaglandins, this specific property of the stomach recently has more frequently been termed *gastroprotection.* Cytoprotection does not actually refer to protection of the entire gastric mucosa. Although gross damage is inapparent when injury is induced after the administration of a protective agent, histologic loss of much of the superficial epithelium is readily observed (12). Gastroprotection does not, therefore, involve the entire mucosa; the protected superficial epithelium quickly reestablishes a structural and electrophysiologic barrier to further injury (13) and deeper necrotic, hemorrhagic lesions are prevented (14).

Mechanisms of Gastric Mucosal Defense

After the important observation of prostaglandin cytoprotection, many research efforts

---

Adapted from Grand Rounds Presentation by Dr. Babyatsky to the Department of Medicine on November 8, 1994. Accepted for publication on October 31, 1995.

Address correspondence and requests for reprints to Dr. Mark W. Babyatsky, Assistant Professor of Medicine, The Mount Sinai School of Medicine, Box 1069, One Gustave L. Levy Place, New York, NY 10029.

were directed toward the development of prostaglandin analogs which might be used in ulcer treatment and also toward understanding the mechanisms that mediate this protection. Prostaglandins were soon shown to increase gastric mucus secretion (15, 16), bicarbonate secretion (17, 18), and mucosal blood flow (19, 20), stabilize mucosal mast cell (21, 22) and lysosomal membranes (23), and maintain endothelial integrity (24). Further, they inhibit free radical production and enzyme release by neutrophils (25, 26) and stimulate rapid restitution of disrupted gastric epithelium (27, 28). Although many of these properties are probably not directly mediated by prostaglandins and many other agents have subsequently been shown to be protective through identical or similar mechanisms, the explosion of research into gastric mucosal defense after Robert's initial observations has allowed the development of a conceptual framework to further understand the complexity of gastric mucosal defense.

Gastric mucosal defense mechanisms can be divided into three layers: preepithelial, epithelial, and subepithelial. At each level, specialized properties allow the gastric epithelium to protect itself against many noxious agents such as acid, pepsin, and ingested toxins. The preepithelial layer lies between the superficial epithelium and the luminal contents. Of these components, the mucus-bicarbonate barrier has received the most attention and has been considered the first line of gastric defense. Gastric mucus, formed in part by noncovalent interactions between large molecular-sized glycoproteins, is a viscoelastic gel which provides a physical barrier as a stable unstirred layer. This gel has the ability to flow and reseal, forming a continuous cover over the mucosa (29). The mucus gel is mostly aqueous and is permeable, at least on its luminal side, to hydrochloric acid and most aqueous solutions, although it is impermeable to pepsin (30). Gastric mucus also provides the structure for the mucus-bicarbonate barrier. Surface epithelial cells release high concentrations of sodium bicarbonate into the mucus gel, sufficient to neutralize much of the hydrochloric acid which back-diffuses into the mucus gel. This bicarbonate barrier creates a significant pH gradient across the thickness of the mucus gel itself, so that the mucus nearest the lumen has a pH of 1 to 2 and that nearest the epithelium is almost neutral (31). As we shall see below, this pH gradient enables *Helicobacter pylori,* a bacterium which traverses the mucus gel, to live in a portion of the mucus gel adjacent to the gastric epithelium. In addition to this bicarbonate gradient, hydrochloric acid released by the parietal cells does not simply diffuse across the mucus gel, but penetrates the mucus in a fine stream at specific points, reaching the lumen by a mechanism known as viscous fingering, a further defense mechanism against acid (32,33).

Even closer to the lumen than the mucus gel lies an additional protective layer. This layer is made up of phospholipids, creating a hydrophobic monolayer at the luminal surface covering the mucus gel (34). These phospholipids resemble pulmonary surfactant in composition, consisting primarily of phosphatidylcholine and phosphatidylethanolamine, and form a hydrophobic layer which prevents water-soluble compounds and acid in the gastric lumen from reaching and damaging the epithelial cells (35). Gastroprotective agents induce the synthesis of the hydrophobic phospholipids, particularly phosphatidylcholine (36). Damaging agents frequently disrupt this protective layer. *H. pylori,* for example, synthesizes phospholipase A2 and ammonium ion, both of which destroy the hydrophobic phospholipid surface.

The epithelial cells themselves also contain specialized properties protective against the unique gastric milieu. The gastric mucosal barrier was originally defined as the relative impermeability of gastric epithelium to passive movement of ions provided by an *electrophysiologic resistance* to back-diffusion of hydrochloric acid. Acute gastric mucosal damage induced by exposure to bile salts, ethanol, salicylates, etc., overwhelms this defense mechanism and leads to a marked increase in ionic permeability and a precipitous decline in transmucosal potential difference (37–39). Although integrity of the barrier was originally described in ionic terms, specialized apical membrane ion transport systems and tight junctions linking adjacent cells provide the anatomic basis for this barrier (40). Another protective property of the specialized gastric epithelium is called restitution, defined as the reestablishment of epithelial cell continuity independent of cell proliferation (41). In this process, superficial defects in the surface epithelium which do not penetrate the basal lamina are rapidly repaired by migration of adjacent viable cells reestablishing epithelial cell continuity within minutes after injury (42, 43). Although initially described as a property of the gastric mucosa and frog skin (44), many epithelial surfaces demonstrate varying capacities for restitution. Several locally expressed growth factors have recently been shown to promote this property of restitution in epithelial cells (45, 46). Such growth factors, either expressed by gastric epithelial cells themselves or through binding receptors found on epithelial cells, play important roles in coordinating epithelial cell responses to preserve normal homeostasis and mucosal integrity in the face of environmental challenges. The fibroblast growth factor family of peptides and platelet-derived growth factor are expressed in the gastric mucosa; basic fibroblast growth factor, although sensitive to acid, can lead to accelerated healing of gastric damage when made resistant to acid by mutation or through binding to sucralfate (47, 48). Epidermal growth factor, produced in saliva (49), and transforming growth factor α, produced locally in the gastric mucosa (50), have been shown to protect against experimental gastric injury (51–53). Other growth factors may also be protective against gastric damage. The trefoil factor family, a recently described group of small, luminally stable peptides, are expressed in mucous cells throughout the gastrointestinal tract (54, 55). Although initially thought to induce cell proliferation, they are highly concentrated in the mucus gel (54) and have recently been shown to protect against gastric damage by ethanol or indomethacin when administered luminally, but not systemically (56).

Also within the epithelial cell layer, endogenous sulfhydryl compounds are found at remarkably high concentrations in the gastric mucosa and appear to protect the gastric epithelium by binding toxic and harmful substances. The most abundant sulfhydryl compound, glutathione, and others such as cysteamine and dimercaptol, are remarkably gastroprotective (57, 58). Further, agents which block sulfhydryl-containing compounds, such as iodoacetamide, lead to severe destruction of the mucosal surface (58). The mechanism by which sulfhydryl compounds protect is unclear, but they are free radical scavengers and have been proposed to mediate prostaglandin gastroprotection, at least in part.

Subepithelial mechanisms of gastrointestinal mucosal defense have received a great deal of attention during the past few years. Recognition of endogenous factors such as nitric oxide, a vasodilator which is gastroprotective (59), and endothelin, a vasoconstrictor which is destructive to the gastric mucosa (60), has underscored the importance of the mucosal and submucosal vascular space in mediating the host response to injury. Essential for the disposal of hydrogen ions and other noxious agents which permeate the mucosal barrier, the gastric intraglandular capillary supply originates from submucosal arterioles which divide into capillaries at the base of the glands, join to a network just beneath the mucus cells, and drain into mucosal venules (61), ensuring the maintenance of a unidirectional blood flow in the mucosal exchange vessels, adequate acid-base balance, and the alkaline tide by transporting bicarbonate from lower regions to more luminal regions of the mucosa. Interactions between the vascular space and the immune system has also been recognized as important in mucosal protection. For example, blocking adherence of neutrophils to vascular endothelial cells by monoclonal antibodies directed against adhesive glycoproteins such as CD11a, CD11b, and CD18 or endothelial intercellular adhesion molecule-1 results in gastroprotection against aspirin, indomethacin, and *H. pylori* (62, 63).

The enteric nervous system also plays a role in gastric mucosal defense; at least two enteric neuropeptides are gastroprotective. When calcitonin gene-related peptide, a neurotransmitter abundantly released by sensory afferent neurons of the stomach, is blocked by capsaicin depletion or specific antibody immunoneutralization, experimental gastric damage is significantly enhanced (64, 65). Parenteral somatostatin, or its synthetic analog octreotide, is gastroprotective against both ethanol and indomethacin gastric injuries (66, 67). Immunoneutralization of somatostatin also leads to aggravated experimental gastric injury. Further, somatostatin gene expression rises dramatically after experimental injury, which suggests a protective role for this factor in gastric mucosal defense (68). Recently, five somatostatin receptors have been cloned (69, 70); binding studies localize these receptors to include the various components of the mucosal immune system (71, 72). As somatostatin inhibits multiple facets of the mucosal immune response (73–75), the mechanism of protection for somatostatin and other immunomodulatory neuropeptides may be through a bidirectional communication between the enteric nervous system and mucosal immune system which has recently been recognized.

After briefly reviewing some of the current foci of research in gastric mucosal defense, I would like to spend the rest of the hour discussing the two recognized clinical correlates of peptic ulcer disease in the 1990s. Although much previous attention has centered on the role of hyperacidity in the pathogenesis of "peptic" ulcer disease, two other factors have emerged as responsible for the vast majority of peptic ulcer disease, nonsteroidal anti-inflammatory drugs (NSAIDs) and *H. pylori*. Peptic ulcer disease directly attributable to high acid states, such as Zollinger-Ellison syndrome, mastocytosis, and basophilic leukemia, probably account for less than 1% of the total population with peptic ulcers.

Nonsteroidal Anti-inflammatory Drugs

The introduction of sodium salicylate in 1875 to treat patients with rheumatic disease was followed by reports of "dyspeptic indigestion" (76). This class of drugs quickly became the most widely used group of drugs with the further development of acetylsalicyclic acid (aspirin) and the ever-expanding number of available nonaspirin NSAIDs (77). Unfortunately, this drug group is also responsible for significant morbidity and mortality. The most frequent adverse effect of NSAIDs is gastric mucosal damage which can lead to gastrointestinal hemorrhage or perforation (78). The risk of hospitalization has been estimated at 1 to 1.5% per year and the risk of death at approximately 0.13% per year from gastrointestinal complications in patients taking NSAIDs (79). Risk factors for increased NSAID-associated injury include age, especially in women (87), past history of ulcer disease (88), the use of more than one NSAID (89), and the concomitant use of corticosteroids (90).

Aspirin and NSAID ingestion induce different types of damage to the gastrointestinal mucosa. Within hours in almost all individuals examined, acute endoscopically visible lesions, consisting of punctate subepithelial petechiae or hemorrhages, can be seen (84, 85). Within days, these early lesions disappear or regress in a process called "adaptive cytoprotection" thought to be secondary to depletion of local factors (prostaglandins) which provide local resistance to injury from the drug itself (86, 87). The earliest lesions are replaced by white-based erosions, most notably in the gastric antrum which wax and wane spontaneously (88). With long-term NSAID use, these erosions become deeper ulcers most commonly in the gastric antrum (89), although NSAIDs cause ulcers throughout the stomach, as well as the small and large intestine (90).

The mechanisms for aspirin and NSAID-induced gastric damage are different for the acute lesions and for chronic ulceration; two modes of injury occur. Acute damage is caused by local irritant effects of the drugs themselves. Microscopic damage is demonstrable by 5 to 10 minutes after oral administration of these agents, before significant cyclooxygenase and prostaglandin inhibition (91, 92). Salicylate, the product of aspirin deacetylation, is toxic to cells and modulates epithelial function. Intraluminally, salicylates result in "barrier" injury, associated with large net cation fluxes, hydrogen back-diffusion, and decrease in transmucosal potential difference. Both apical and basolateral plasma membrane resistances and mucosal content of ATP is reduced with consequent stimulation of sodium transport and proton leaks from surface epithelial cells (93).

In addition, aspirin disrupts gastric mucosal defense, independent of direct salicylate toxicity, through the inhibition of cyclooxygenase and prostaglandin synthesis. In fact, sodium salicylate, which does not inhibit prostaglandins, does not result in chronic ulceration (94). In 1971, Vane (95) observed that aspirin and other NSAIDs inhibit the synthesis of prostaglandins and postulated that prostaglandin inhibition was the mechanism for the gastric toxicity of these compounds. The doses of NSAIDs required to inhibit cyclooxygenase, the prostaglandin-synthesizing enzyme, occur within the therapeutic range of these drugs (96). Recognition of the role of prostaglandin inhibition in the pathogenesis of NSAID gastroduodenal ulceration has led to the development of the first prostaglandin analog, misoprostol, which has been used successfully in both the treatment and prevention of NSAID-associated gastric and duodenal ulcers (97–99). Other prostaglandin analogs are currently under development. Although peptic ulceration associated with NSAIDs can spontaneously regress even with continued used of the NSAID (100), histamine receptor-2 blockers (101, 102), sucralfate (103), and proton-pump inhibitors (104), in addition to misoprostol, all lead to increased healing with continued used of the NSAID. However, no agent has yet proved equal or superior to misoprostol in prevention of NSAID-associated gastric or duodenal ulcers (105).

HELICOBACTER PYLORI

In 1676, Anton Van Leeuwenhoek published the following statement, "In the year 1675, I discovered living creatures in rain water. This invited me to view this water with great attention" (106). However, spiral organisms were not noted in gastric mucosa until two centuries later, and even after these organisms were described, they were dismissed as irrelevant for yet another century. In 1892, Bizzozero (107) noted spiral-shaped organisms in the stomachs of cats and dogs. In 1906, Krienitz (108) reported similar spiral-shaped bacteria in human autopsies in anatomic association with gastric cancer. The ensuing 30 years brought scattered reports of these organisms in the stomachs of patients with benign peptic ulcers (109–111). An extensive autopsy series in 1938 demonstrated a 43% prevalence for spiral organisms in the human stomach (112), but the author noted no relationship between the presence of organisms and gastric pathology, leading him to conclude that bacteria seen associated with the gastric mucosa represented bacterial contaminants introduced by swallowing of oral contents. This hypothesis received further support with the publication of a series of 1000 patients by Palmer (113). Although gastric organisms received little further attention until their rediscovery in the 1980s, antibiotics were noted to reduce gastric urease levels (114), although the source of gastric urease, at the time, was felt to be the gastric epithelium.

In 1982, a medical intern, Barry Marshall, M.D., on rotation with a pathologist, J. Robin Warren, in Perth, Australia noted the presence of spiral-shaped organisms on gastric endoscopic biopsies associated with the pathologic conditions of gastritis, peptic ulcers, and gastric carcinomas (115, 116). Overcoming initial rampant skepticism, Marshall was eventually successful in culturing the organism and filling Koch's postulates in its associations with gastric disease (117). The organism is a unipolar, multiflagellate, curved or gently spiral microaerophilic Gram-negative rod which penetrates the mucus layer overlying the gastric epithelium where the pH is close to neutral (106). The S-shape of the organism allows for corkscrew-like movements and penetration of the gastric mucus layer; microaerophilic metabolism ensures survival in the semipermeable mucus gel; urea forms a cloud around the organism maintaining a neutral pH during passage through an acidic milieu.

The most common chronic bacterial infection in the world, *H. pylori* infects approximately one third of the world's population (118). The prevalence of *H. pylori* infection is highest in underdeveloped countries; childhood infection is common and 70% of adults older than 25 years of age are infected (119). In developed countries, infection is uncommon in childhood but occurs in 50% of adults older than 60 years of age. *H. pylori* infection also correlates with socioeconomic status: with rising socioeconomic status, prevalence of infection has declined, underlying decreasing rates of gastritis and peptic ulcer disease noted in developed countries throughout the 20th century (120, 121).

The mechanism of transmission of *H. pylori* infection remains unclear. There appears to be direct person-to-person transmission. However, although most epidemiologic evidence supports fecal-oral transmission (122, 123), oral-oral transmission has not been ruled out conclusively and evidence of *H. pylori* in dental plaque has been found by culture (124, 125) and by reverse polymerase chain reaction (126). Although man is, by far, the most common host, nonhuman reservoirs of *H. pylori* have recently been detected in cats, primates, and possibly pigs (127).

In 1995, there remains no question that *H. pylori* is *the cause* of chronic superficial type B gastritis. Many studies have demonstrated a clear relationship with the presence of histologic gastritis; multiple studies have even correlated the density of *H. pylori* infection with the severity of inflammation (128–130). Further, oral challenge in human volunteers (117, 131) and animal models (132–134) have demonstrated consistent superficial gastritis. All patients infected with *H. pylori* demonstrate antral gastritis; eradication results in resolution of the gastritis (135).

Both duodenal and gastric ulcer disease are closely associated with *H. pylori* infection; *H. pylori* infection can be clearly demonstrated in 90 to 100% of patients with duodenal ulcers (116, 136) and 60 to 100% of patients with gastric ulcers (116, 136, 137). Prospective studies soon solidified the association by showing a marked reduction in ulcer recurrence after eradication of *H. pylori.* Graham et al. (138) examined 109 patients with *H. pylori* infection and recent peptic ulcer disease who were treated either with ranitidine alone or with ranitidine plus tetracycline, metronidazole, and bismuth. Two years after treatment, the recurrence rate for duodenal ulcer by endoscopic examination is 12% for the group treated with antibiotics compared with 95% for the group treated with ranitidine alone. The results for gastric ulcers were similar: 13% for antibiotics and 74% for ranitidine alone. Another prospective study by Hentshel et al. (139) of 104 Austrian patients with *H. pylori* infection and recurrent duodenal ulcers who were treated with either amoxacillin, metronidazole, and ranitidine or ranitidine alone. At 1 year, duodenal ulcers were present endoscopically in 8% of the patients given antibiotics with ranitidine compared with 86% given ranitidine alone. These studies suggest a causal role for *H. pylori* in recurrent duodenal ulcer disease.

Although the presence of *H. pylori* appears to be necessary for the development of almost all non-NSAID peptic ulcers, its presence does not appear to be sufficient for ulcer formation. In fact, only 1.7 to 20% of patients with histologic or serologic evidence of *H. pylori* infection will ever develop gastric or duodenal ulcers (136, 140, 141). Other factors, such as virulence factors intrinsic to the organism, or other extrinsic factors which may further disturb gastric mucosal defenses, must determine which patients infected with *H.*

pylori will develop peptic ulcer disease. Strains of *H. pylori* are genetically heterogeneous; cytotoxins produced by a subset of these strains may be important in ulcerogenesis. Both a 120-kDa vacuolating cytotoxin (142, 143) and the cagA gene product (144, 145) have been associated with peptic ulcers. However, no bacterial products have yet been identified which accurately predict those infected patients who will develop ulcers (146). Host factors which increase the risk of developing ulcers in infected patients include cigarette smoking, stress, and use of NSAIDs (136, 147). Concordance rates for peptic ulcers are higher for monozygotic than for dizygotic twins, although it is unclear whether this is caused by an increased genetic susceptibility to *H. pylori* infection (121, 148). Other genetic factors which may play a role in ulcer development include blood group antigens (149, 150) and HLA haplotypes (151).

The precise mechanisms by which *H. pylori* infection leads to peptic ulceration remains unclear. Most physiologic alterations long associated with peptic ulcer disease are reproduced by *H. pylori* including increases in gastrin secretion (152–154), inhibition of antral somatostatin release (155, 156), and decreased duodenal bicarbonate secretion (157). The net result of these changes is increased acid output and decreased acid neutralization leading to gastric metaplasia of the duodenum and the creation of an environment conducive to *H. pylori* colonization (158). However, the steps leading from subsequent duodenitis to actual ulceration in a subset of patients remains unclear and most likely relate to host factors or *H. pylori* virulence factors as described above. Similarly, the sequence of events from gastritis to the formation of gastric ulcers remains unclear.

Diagnosis of *H. pylori* infection can be performed by invasive or noninvasive means. Esophagogastroduodenoscopy is indicated for symptomatic patients to determine whether ulcers are present. At the time of endoscopy, mucosal biopsy specimens should be obtained for all patients with gastric or duodenal ulcers. The presence of *H. pylori* can then be determined by culture, histologic examination, or by a rapid urease test such as the CLO test (159). Because culture is relatively insensitive and histologic examination costly, the CLO test can be performed initially while saving some mucosal tissue in formalin. If the results of the CLO test are negative, the specimen can then be sent for pathology to confirm the absence of *H. pylori*. Noninvasive methods of detection include serology (160) and breath tests (161, 162), based on the breakdown of C-13 or C-14 labeled ingested urea by *H. pylori* urease. As the sensitivity and specificity of serologic testing with IgA or IgG antibodies approaches that of histologic detection (163), this has become the diagnostic choice for initial determination of infection in patients with a history of documented history of peptic ulcer disease in whom endoscopic examination is either not clearly indicated or contraindicated by coexisting illness. However, because antibody titers take months to decrease after eradication and do not decrease in all patients with documented eradication (164), serologic testing has not been useful for demonstrating eradication after treatment. For confirmation of *H. pylori* eradication, breath testing, after availability increases and the Food and Drug Administration approves its use, will probably become the method of choice.

Treatment for patients infected with *H. pylori* remains challenging. First, the indications for treatment have undergone several important changes and will likely be changed again in the near future. A National Institutes of Health Consensus Development Conference recently recommended that all patients with documented ulcers, gastric or duodenal, who are infected with *H. pylori* should be treated to eradicate the organism, including those who have taken concurrent NSAIDs and those with a remote history of peptic ulcer disease (165). Treatment of patients with nonulcer dyspepsia remains more confusing. No clear evidence yet exists that eradication of *H. pylori* in patients with dyspeptic symptoms in the absence of ulceration leads to a significant improvement in the clinical course of the illness (166). Because the symptoms of dyspepsia and peptic ulcer disease are often indistinguishable and dyspepsia is a very common group of symptoms, only patients with documented peptic ulcers should currently undergo eradication treatment. However, should prospective, randomized trials identify subsets of patients with dyspeptic symptoms who benefit from such therapy, these recommendations will change. Further, if the link between *H. pylori* and gastric cancer continues to be solidified, recommendations for therapy may include an even broader population of infected patients.

A second challenge of therapy is the treatment itself. Although many antimicrobial agents demonstrate *in vitro* susceptibility, *in vivo* activity of many agents does not correlate well, and resistance to antibiotics frequently develops during the course of therapy, particularly if an agent is used alone (167). Successful treatment requires combination therapy with more than one antimicrobial agent (168, 169). Although many combination regimens are effective in eradication of *H. pylori*, none are effective 100% of the time and most have significant side effects. Other challenges of therapy include the relative inability of antibiotic penetrance into the mucus gel where *H. pylori* reside and the inactivation of a number of the effective antibiotics at the acid pH of the stomach. The latter may explain the increased efficacy in eradication of *H. pylori* with the addition of acid blockers to many of the antimicrobial regimens. At present, many regimens achieve eradication rates of approximately 90% (146) and the specific choice of regimen seems to be rather arbitrary: cost of the regimen and side effect profiles should clearly be considered in the selection of a specific regimen. Although initial reports of using a two-drug regimen of high-dose omeprazole and amoxacillin were promising (170), more recent reports have failed to support this enthusiasm (171, 172). Currently, the most frequently recommended regimen consists of bismuth subsalicylate (Pepto Bismol), metronidazole, and tetracycline or amoxacillin for 2 weeks because of its cost and efficacy (168, 169). The use of antimicrobial agents plus a histamine receptor antagonist or proton pump inhibitor demonstrates similar efficacy but at much higher cost (146). Both omeprazole and lansoprazole are bactericidal for *H. pylori in vitro,* but are not capable of eradication of the organism *in vivo* (173, 174). As the regimens mentioned remain less than optimal because of cost, side effects, and the number of agents needed with which to comply, the future challenge will be to find a regimen that is easier to take at the same or higher eradication rate. Once *H. pylori* is eradicated, the reinfection rate is low (175), indicating that the host immune response may be effective in preventing recurrence of peptic ulcer disease.

In addition to improved therapy, long-term strategies for dealing with the consequences of infection with *H. pylori* will focus on prevention. The focus of this strategy will be to develop a safe and effective vaccine. Recent developments including a new mouse model infected with *H. pylori* itself (132) and demonstration that oral immunization with killed *Helicobacter felis* plus cholera toxin can prevent experimental *H. felis* infection (176, 177), suggest that a vaccine may be forthcoming in the near future. An ideal vaccine would be an oral vaccine which could be administered in infancy on a schedule similar to other vaccination programs because the incidence of *H. pylori* infection is highest in infancy, particularly in underdeveloped regions. The development of a vaccine may have a dramatic impact on both non-NSAID peptic ulcer disease and gastric cancer, which are common and serious health problems throughout the world.

In short, throughout much of the twentieth century, acid was believed to play a central role in the pathogenesis of peptic ulcer disease. More recently, prostaglandins and a rapidly expanding list of agents have been demonstrated to be important in gastric mucosal defense. The major development in the understanding of this disease has been the recognition of a bacterium responsible for the vast majority of peptic ulcers. Further research into gastric mucosal defense mechanisms, including strategies to prevent NSAID-related injury, and better treatment and prevention of *H. pylori* infection, may relegate peptic ulcer disease to a category of rare disorders. The dramatic recent emergence of the factors responsible for peptic ulcer disease has demonstrated the importance of openness toward new paradigms as we discard outdated paradigms which have temporarily blinded further progress in the understanding of the pathogenesis of human disease.

References

1. Kurata JH, Haile BM. Epidemiology of peptic ulcer disease. Clin Gastroenterol 1984; 13:289–292.
2. Isenberg JI, McQuaid KR, Laine L, Walsh JH. Acid-peptic disorders. In: Yamada T, editor. Textbook of gastroenterology. Philadelphia: Lippincott, 1995: 1347–1430.
3. Schwartz K. Uber penetrierende magen-und jejunal geshcwure. Beitr Klin Chir 1910; 76:96–128.
4. Black JW, Duncan WAM, Durant GJ, et al. Definition and antagonism of histamine H2-receptors. Nature 1972; 236:385–388.
5. Bardhan KD, Thompson M, Bose K, et al. Combined anti-muscarinic and H2-receptor blockade in the healing of refractory duodenal ulcer. A double blind study. Gut 1987; 28:1505–1512.
6. Blair AJ III, Feldman M, Barnett C, et al. Detailed comparison of basal and food-stimulated gastric acid secretion rates and serum gastrin concentrations in duodenal ulcer patients and normal subjects. J Clin Invest 1987; 79:582–590.
7. Samloff IM. Peptic ulcer: the many proteinases of aggression. Gastroenterology 1989; 96(Suppl):586–595.
8. Soll AH. Pathogenesis of peptic ulcer and implications for therapy. N Engl J Med 1990; 322:909–917.
9. Robert AR, Schultz JR, Nezamis JE, et al. Gastric antisecretory and antiulcer properties of PGE2, 15-methyl PGE2 and 16,16-dimethyl PGE2 intravenous, oral and intrajejunal administration. Gastroenterology 1976; 70:359–368.
10. Robert A, Nezamis JE, Phillips JP. Effect of prostaglandin E1 on gastric secretion and ulcer formation in the rat. Gastroenterology 1968; 55:481–487.
11. Robert Andre: Cytoprotection by prostaglandins. Gastroenterology 1979; 77:761–767.
12. Lacy ER, Ito S. Microscopic analysis of ethanol damage to rat gastric mucosa after treatment with prostaglandins. Gastroenterology 1982; 83:619–625.

13. Silen W. What is cytoprotection of the gastric mucosa? Gastroenterology 1988; 94:232–235.
14. Konturek SJ. Mechanisms of gastroprotection. Scand J Gastroenterol 1990; 25(Suppl 174):15–28.
15. Bickel M, Kauffman GL. Gastric gel mucus thickness: Effect of distention, 16,16-dimethyl prostaglandin E2, and carbenoxolone. Gastroenterology 1981; 80:770–778.
16. McQueen S, Hutton D, Allen A, et al. Gastric and duodenal surface mucus gel thickness in rat: effects of prostaglandins and damaging agents. Am J Physiol 1983; 245:388–395.
17. Bolton JP, Cohen MM. Stimulation of non-parietal cell secretion in canine Heidenhain pouches by 16,16-dimethyl prostaglandin E2. Digestion 1978; 17:291–298.
18. Kauffman GL, Reeve JJ, Grossman MI. Gastric bicarbonate secretion: effect of topical and intravenous 16,16-dimethyl prostaglandin E2. Am J Physiol 1980; 239:44–49.
19. Gerber JG, Neis AS. Canine gastric mucosal vasodilation with prostaglandins and histamine analogs. Dig Dis Sci 1982; 27:870–878.
20. Main IHM, Whittle BJR. The effects of E and A prostaglandins on gastric mucosal blood flow and acid secretion in the rat. Br J Pharmacol 1973; 49:428–435.
21. Raud J. Vasodilatation and inhibition of mediator release represent two distinct mechanisms for prostaglandin modulation of acute mast cell-dependent inflammation. Br J Pharmacol 1990; 90:449–454.
22. Hogaboam CM, Bissonnette EY, Chin BC, et al. Prostaglandins inhibit inflammatory mediator release from rat mast cells. Gastroenterology 1993; 104:122–129.
23. Ferguson WW, Demonds AW, Starling JR, et al. Protective effect of prostaglandin E1 (PGE1) on lysosomal enzyme release in serotonin-induced gastric ulceration. Ann Surg 1973; 177:643.
24. Guth PH, Paulsen G, Nagata H. Histologic and microcirculatory changes in alcohol-induced gastric lesions in the rat: effect of prostaglandin cytoprotection. Gastroenterology 1984; 87:1083–1090.
25. Gryglewsin RJ, Szczklik A, Wandzilak M. The effect of six prostaglandins, prostacyclin and iloprost on generation of superoxide anions by human polymorphonuclear leukocytes stimulated by zymozan or formylleucyl-phenylalanine. Biochem Pharmacol 1987; 36: 4209–4218.
26. Ham EA, Soderman DD, Zanetti ME, et al. Inhibition by prostaglandins of leukotriene B_4 release from activated neutrophils. Proc Natl Acad Sci USA 1983; 80: 4349–4353.
27. Svanes K, Critchlow J, Takewchi K, et al. Factors influencing reconstitution of frog gastric mucosa: role of prostaglandins. In: Allen A, Flemstrom G, Garner A, Silen W, Turnberg LA, editors. Mechanisms of mucosal protection in the upper gastrointestinal tract. Raven Press: New York, 1984: 33–38.
28. Ito S, Lacy ER. Morphology of rat gastric mucosal damage, defense and restitution in the presence of luminal ethanol. Gastroenterology 1985; 88:250–260.
29. Allen A. The structure and function of gastrointestinal mucus. In: Harmon JW, editor. Basic mechanisms of gastrointestinal mucosal cell injury and protection. Baltimore: Williams and Wilkins, 1981: 351–367.
30. Silberberg A, Meyer FA, Gilboa A, Gelman RE. Function and properties of epithelial mucus. In: Elstein M, Parke DV, editors. Mucus in health and disease. New York: Plenum Press, 1977: 171–180.
31. Bahari HMM, Ross IN, Turnberg LA. Demonstration of a pH gradient across the mucus layer on the surface of human gastric mucosa *in vitro*. Gut 1982; 23:513–516.
32. Bhaskar KR, Garik P, Turner BS, et al. Viscous fingering of HCl through gastric mucin. Nature 1992; 360: 458–461.
33. Fabry TL. How parietal secretion crosses the gastric mucus without being neutralized. Gastroenterology 1990; 98:A42.
34. Hills BA, Butler BD, Lichtenberger LM. Gastric mucosal barrier: hydrophobic lining to the lumen of the stomach. Am J Physiol 1983; 244:G561–G568.
35. Butler BD, Lichtenberger LM, Hills BA. Distribution of surfactants in the canine gastrointestinal tract and their ability to lubricate. Am J Physiol 1983; 244: G645–G651.
36. Lugea A, Mourelle M, Guarner F, et al. Phosphatidylcholines as mediators of adaptive cytoprotection of the rat duodenum. Gastroenterology 1994; 107:720–727.
37. Davenport HW. Gastric mucosal injury by fatty and acetylsalicylic acid. Gastroenterology 1965; 46:245–253.
38. Davenport HW. Ethanol damage to canine oxyntic glandular mucosa. Proc Soc Exp Biol Med 1967; 126:657–662.
39. Davenport HW. Salicylate damage to the gastric mucosal barrier. N Engl J Med 1967; 276:1307–1312.
40. Allen A, Flemstrom G, Garner A, Kivilaakso E. Gastroduodenal mucosal protection. Physiol Rev 1993; 73: 823–857.
41. Silen W, Ito S. Mechanisms for rapid reepithelialization of the gastric mucosal surface. Annu Rev Physiol 1985; 47:217–229.
42. Lacy ER, Ito S. Ethanol-induced insult to superficial rat gastric epithelium: a study of damage and rapid repair. In: Allen A, editor. Mechanisms of mucosal protection in the upper gastrointestinal tract. New York: Raven Press, 1984: 49–56.
43. Rutten MJ, Ito S. Morphology and electrophysiology of guinea pig gastric mucosal repair in vitro. Am J Physiol 1983; 244:G171–G182.
44. Svanes K, Ito S, Takeuchi K, Silen W. Restitution of the surface epithelium of the in vitro frog gastric mucosa after damage with hyperosmolar sodium chloride. Morphological and physiological characteristics. Gastroenterology 1982; 82:1409–1426.
45. Dignass A, Lynch-Devaney K, Kindon H, et al. Trefoil peptides promote epithelial migration through a transforming growth factor β-independent pathway. J Clin Invest 1994; 94:376–383.
46. Dignass A, Podolsky DK. Cytokine modulation of intestinal epithelial cell restitution: central role of transforming growth factor β. Gastroenterology 1993; 105: 1323–1332.
47. Folkman J, Szabo S, Stovroff M, et al. Duodenal ulcer: discovery of a new mechanism and development of angiogenic therapy that accelerates healing. Ann Surg 1991; 214:414–426.
48. Folkman J, Szabo S, Vattay P, et al. Effect of orally administered bFGF on healing of chronic duodenal ulcers, gastric secretion and acute mucosal lesions. Gastroenterology 1990; 98:A45.
49. Heitz PU, Kasper M, Van Noorden S, et al. Immunohistochemical localization of urogastrone to human duodenal and submandibular glands. Gut 1978; 19:408–413.

50. Beauchamp RD, Barnard JA, McCutchen CM, et al. Localization of transforming growth factor alpha and its receptor in gastric mucosal cells: implications for a regulatory role in acid secretion and mucosal renewal. J Clin Invest 1989; 84:1017–1023.
51. Tepperman BL, Soper BD. Effect of epidermal growth factor, transforming growth factor a and nerve growth factor on gastric mucosal integrity and microcirculation in the rat. Regul Pept 1994; 50:13–21.
52. Olsen PS. Role of epidermal growth factor in gastroduodenal mucosal protection. J Clin Gastroenterol 1988; 10(Suppl 1):S146–S151.
53. Romano M, Polk WH, Awad JH, et al. Transforming growth factor alpha protection against drug-induced injury to the rat mucosa in vivo. J Clin Invest 1992; 90:2409–2421.
54. Jeffrey G, Oates P, Wang T, et al. Spasmolytic polypeptide: a trefoil peptide secreted by rat gastric mucosal cells. Gastroenterology 1994; 106:336–345.
55. Sueomori S, Lynch-Devaney K, Podolsky DK. Identification and characterization of rat intestinal trefoil factor: tissue and cell-specific member of the trefoil protein family. Proc Natl Acad Sci USA 1991; 88:11017–11021.
56. Babyatsky MW, de Beaumont M, Thim L, Podolsky DK. Oral trefoil peptides protect against ethanol and indomethacin-induced gastric injury in rats. Gastroenterology (in press).
57. Szelenyi I, Brune K. Possible role of sulfhydryls in mucosal protection induced by aluminum hydroxide. Dig Dis Sci 1986; 31:1207–1210.
58. Szabo S, Trier JS, Frankel PW. Sulfhydryl compounds may mediate gastric cytoprotection. Science 1981; 214: 200–202.
59. Whittle BJR, Lopez-Belmonte J, Moncada S. Regulation of gastric mucosal integrity by endogenous nitric oxide: interactions with prostanoids and sensory neuropeptides in the rat. Br J Pharmacol 1990; 99:607–611.
60. Morales RE, Johnson BR, Szabo S. Endothelin induces vascular and mucosal lesions, enhances the injury by HCl/ethanol, and the antibody exerts gastroprotection. FASEB J 1992; 6:2354–2360.
61. Gannon B, Browning J, O'Brien P, Rogers P. Mucosal microvascular architecture of the fundus and body of human stomach. Gastroenterology 1984; 86:866–875.
62. Yoshida N, Granger DN, Evans DJ Jr, et al. Mechanisms involved in *Helicobacter pylori*-induced inflammation. Gastroenterology 1993; 105:1431–1440.
63. Yoshida N, Takemura T, Granger DN, et al. Molecular determinants of aspirin-induced neutrophil adherence to endothelial cells. Gastroenterology 1993; 105:715–724.
64. Forster ER, Dockray GJ. The role of calcitonin gene-related peptide in gastric mucosal protection in the rat. Exp Physiol 1991; 76:623–626.
65. Pabst MA, Schoninkle E, Holzer P. Ablation of capsaicin sensitive afferent nerves impairs defense but not rapid repair of rat gastric mucosa. Gut 1993; 34:897–903.
66. Diel F, Szabo S. Dose-dependent effects of linear and cyclic somatostatin on ethanol-induced gastric erosions: the role of mast cells and increased vascular permeability in the rat. Regul Pept 1986; 13:235–243.
67. Karmeli F, Eliakim R, Okon E, Rachmilewitz D. Somatostatin effectively prevents ethanol- and NSAID-induced gastric mucosal damage in rats. Dig Dis Sci 1994; 39:617–625.
68. Babyatsky MW, Brand SJ, Karmeli F, et al. Enhanced somatostatin expression modulates gastric injury and experimental colitis. Gastroenterology 1993; 104: A663.
69. Yamada Y, Post S, Wang K, et al. Cloning and functional characterization of a family of human and mouse somatostatin receptors expressed in brain, gastrointestinal tract, and kidney. Proc Natl Acad Sci USA 1993; 90:4196–4200.
70. Bell GI, Reisine T. Molecular biology of somatostatin receptors. Trends Neurosci 1993; 16:34–38.
71. Lewin MJ. The somatostatin receptor in the GI tract. Annu Rev Physiol 1992; 54:455–468.
72. Reubi JC. Somatostatin receptors in the gastrointestinal tract in health and disease. Yale J Biol Med 1992; 65:493–503.
73. Scicchitano R, Bienenstock J, Stanisz AM. The differential effect with time of neuropeptides on the proliferative responses of murine Peyer's patches and splenic lymphocytes. Brain Behav Immunol 1987; 1:231–237.
74. Payan DG, Hess CA, Goetzl EJ. Inhibition by somatostatin of the proliferation of T-lymphocytes and Molt-4 lymphoblasts. Cell Immunol 1984; 84:433–438.
75. Van Hugen PM, Krenning EP, Kwekkeboom DJ, et al. Somatostatin and the immune and haematopoietic system: a review. Eur J Clin Invest 1994; 24:91–99.
76. Schoen RT, Vender RJ. Mechanisms of nonsteroidal anti-inflammatory drug-induced gastric damage. Am J Med 1989; 86:449–458.
77. Baum C, Kennedy DL, Forbes MB. Utilization of nonsteroidal antiinflammatory drugs. Arthritis Rheum 1985; 28:686–692.
78. Semble EL, Wu WC. Antiinflammatory drugs and gastric mucosal damage. Semin Arthritis Rheum 1987; 16:271–286.
79. Fries JF, Williams CA, Bloch DA. Nonsteroidal anti-inflammatory drug-associated gastropathy: incidence and risk factor models. Am J Med 1991; 91:213–222.
80. Griffin MR, Piper JM, Daugherty JR, et al. Nonsteroidal anti-inflammatory drug use increased risk for peptic ulcer disease in elderly persons. Ann Intern Med 1991; 114:257–263.
81. Nonsteroidal anti-inflammatory drugs and serious gastrointestinal adverse reaction. Br Med J 1986; 292: 614–623.
82. Langman MJS. Epidemiologic evidence on the association between peptic ulceration and anti-inflammatory drug use. Gastroenterology 1989; 96:640–646.
83. Piper JM, Daugherty WA, Griffin MR. Corticosteroid use and peptic ulcer disease: role of nonsteroidal anti-inflammatory drugs. Ann Intern Med 1991; 114:735–740.
84. Levi S, Shaw-Smith C. Non-steroidal anti-inflammatory drugs: how do they damage the gut? Br J Rheum 1994; 33:605–612.
85. Graham DY, Smith JL, Dobbs SM. Gastric adaptation occurs with aspirin administration in man. Dig Dis Sci 1983; 28:1–6.
86. Chaudhury TK, Robert A. Prevention by mild irritants of gastric necrosis produced in rats by sodium taurochholate. Dig Dis Sci 1980; 25:830–836.
87. Lacy ER, Cowart KS, Hund P. Effects of chronic superficial injury on the rat gastric mucosa. Gastroenterology 1992; 103:1179–1191.
88. Larkai EN, Smith JL, Lidsky MD, Graham DY. Gastroduodenal mucosa and dyspeptic symptoms in arthritic patients during chronic nonsteroidal anti-in-

flammatory drug use. Am J Gastroenterol 1987; 82:1153–1158.
89. Hollander D. Gastrointestinal complications of nonsteroidal anti-inflammatory drugs: prophylactic and therapeutic strategies. Am J Med 1994; 96:274–281.
90. Bjarnason I, Hayllar J, MacPherson A, Russell AS. Side effects of NSAIDs on the small and large intestine in humans. Gastroenterology 1993; 104:1832–1847.
91. Baskin WN, Ivey KJ, Krausz WJ, et al. Aspirin-induced ultrastructural changes in human gastric mucosa. Ann Int Med 1976; 85:229–303.
92. Whittle BJR. Temporal relationship between cyclooxygenase inhibition, as measured by prostacyclin biosynthesis, and the gastrointestinal damage induced by indomethacin in the rat. Gastroenterology 1981; 80:94–98.
93. Kauffman G. Aspirin-induced gastric mucosal injury: lessons learned from animal models. Gastroenterology 1989; 96:606–614.
94. Levi S, Shaw-Smi C. Non-steroidal anti-inflammatory drugs: how do they damage the gut? Br J Rheumatol 1994; 33:605–612.
95. Vane JR. Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs. Nature New Biol 1971; 231:232–235.
96. Whittle BJR, Vane JR. A biochemical basis for the gastrointestinal toxicity of nonsteroidal antirheumatoid drugs. Arch Toxicol 1984; 7:315–322.
97. Graham DY, Agrawal NM, Roth SH. Prevention of NSAID-induced gastric ulcer with misoprostol: multicenter, double-blind, placebo-controlled trial. Lancet 1988; 2:1277–1280.
98. McCarthy DM. Misoprostol. Drug Ther 1989; 19:59–69.
99. Lanza FL, Aspinall RL, Swabb EA, et al. Double-blind, placebo-controlled endoscopic comparison of the mucosal protective effects of misoprostol versus cimetidine on tolmetin-induced mucosal injury to the stomach and duodenum. Gastroenterology 1988; 95:289–294.
100. O'Laughlin JC, Silvoso GR, Ivey KJ, Healing of aspirin-associated peptic ulcer disease despite continued salicylate ingestion. Arch Intern Med 1981; 141:781–783.
101. Berkowitz JM, Adler SN, Sharp JT, Warner CW. Reduction of aspirin-induced gastroduodenal mucosal damage with ranitidine. J Clin Gastroenterol 1986; 8:377–380.
102. Robinson MG, Griffin JW Jr, Bowers J, et al. Effect of ranitidine on gastroduodenal mucosal damage induced by nonsteroidal antiinflammatory drugs. Dig Dis Sci 1989; 34:424–428.
103. Caldwell JR, Roth SW, Wu WC, et al. Sucralfate treatment of nonsteroidal anti-inflammatory drug-induced gastrointestinal symptoms and mucosal damage. Am J Med 1987; 83:74–82.
104. McCarthy DM. Nonsteroidal antiinflammatory drug-induced ulcers: management by traditional therapies. Gastroenterology 1989; 96(Suppl):662–674.
105. Lanza F, Peace K, Gustitus L, et al. A blinded endoscopic comparative study of misoprostol versus sucralfate and placebo in the prevention of aspirin-induced gastric and duodenal ulceration. Am J Gastroenterol 1988; 83:143–146.
106. Godwin CS, Worsley BW. Microbiology of *Helicobacter pylori*. Gastroenterol Clin North Am 1993; 22:5–18.
107. Bizzozero G. Uber die schlauchformigen drusen des magendarmkanals und die beziehungen ihres epithels zu dem oberfachenepithel der schleimhaut. Arch Mikrobiol Anat 1892; 42:82–152.
108. Krienitz W. Ueber das auftreten von spirochaten verschiedener form im mageninhalt bei carcinoma ventriculi. Deutsch Med Wochenschr 1906; 28:872.
109. Rosenow EC, Sanford AH. The bacteriology of ulcer of the stomach and duodenum in man. J Infect Dis 1915; 17:219–226.
110. Appelmans R, Vassiliadis P. Etude sur la flore microbienne des ulceres gastroduodenaux et des cancers gastriques. Rev Belge Sci Med 1932; 4:198–203.
111. Freedburg AS, Barron LE. The presence of spirochetes in human gastric mucosa. Am J Dig Dis 1940; 7:443–445.
112. Doenges JL. Spirochetes in the gastric glands of Macacus rhesus and humans without definite history of related disease. Proc Soc Exp Med Biol 1938; 38:536–538.
113. Palmer ED. Investigation of the gastric mucosa spirochetes of the human. Gastroenterology 1954; 27:218–220.
114. Lieber CS, Lefevre A. Ammonia as a source of gastric hypoacidity in patients with uremia. J Clin Invest 1959; 38:1271–1277.
115. Marshall BJ. Unidentified curved bacilla in gastric epithelium in active chronic gastritis. Lancet 1983; 1:1273–1275.
116. Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet 1984; 1:1311–1315.
117. Marshall BJ, Armstrong JA, McGechie DB, Glancy RJ. Attempt to fulfill Koch's postulates for pyloric Campylobacter. Med J Aust 1985; 142:436–439.
118. Taylor DN, Blasser MJ. The epidemiology of *Helicobacter pylori* infections. Epidemiol Rev 1991; 13:42–59.
119. Graham D, Go M. *Helicobacter pylori:* current status. Gastroenterology 1993; 105:279–282.
120. Haruma K, Kawaguchi H, Kohmoto K, et al. Reduced incidence of *Helicobacter pylori* in Japan during the last 10 years. Gastroenterology 1994; 106:A91.
121. Malaty HM, Engstrand L, Pedersen NL, Graham DY. *Helicobacter pylori* infection: genetic and environmental influences. Ann Intern Med 1994; 120:982–986.
122. Megraud F. Transmission of *Helicobacter pylori:* fecal-oral versus oral-oral route. Aliment Pharmacol Ther 1995; 9(Suppl 2):85–91.
123. Thomas JE, Gibson GR, Darboe MK, et al. Isolation of *Helicobacter pylori* from human faeces. Lancet 1992; 340:1194–1195.
124. Krajden S, Fuska M, Anderson J, et al. Examination of human stomach biopsies, saliva and dental plaque for *Campylobacter pylori.* J Clin Microbiol 1989; 27:1397–1398.
125. Shames B, Krajden S, Fuksa M, et al. Evidence for the occurrence of the same strain of *Campylobacter pylori* in the stomach and dental plaque. J Clin Microbiol 1989; 27:2849–2850.
126. Nguyen AH, Engstrand L, Genta RM, et al. Detection of *Helicobacter pylori* in dental plaque by reverse transcription-polymerase chain reaction. J Clin Microbiol 1993; 31:783–787.
127. Fox JG. Non-human reservoirs of *Helicobacter pylori.* Aliment Pharmacol Ther 1995; 9(Suppl 2):93–103.
128. Bayerdorffer E, Lehn N, Hatz R, et al. Difference in expression in *Helicobacter pylori* gastritis in antrum and body. Gastroenterology 1992; 102:1575–1582.
129. Alam K, Schubert TT, Bologna SD, Ma CK. Increased density of *Helicobacter pylori* on antral biopsy is associated with severity of acute and chronic inflammation

and likelihood of duodenal ulceration. Am J Gastroenterol 1992; 87:424–428.
130. Satoh K, Kimura K, Yoshida Y, et al. A topographical relationship between *Helicobacter pylori* and gastritis: quantitative assessment of *Helicobacter pylori* in the gastric mucosa. Am J Gastroenterol 1991; 86:285–291.
131. Morris A, Nicholson G. Ingestion of *Campylobacter pyloridis* causes gastritis and raises fasting gastric pH. Am J Gastroenterol 1987; 82:192–197.
132. Marchetti M, Arico B, Burroni D, et al. Development of a mouse model of *Helicobacter pylori* infection that mimics human disease. Science 1995; 267:1655–1658.
133. Fukuda Y, Tamura K, Yamamoto I, et al. Inoculation of rhesus monkeys with human *Helicobacter pylori*: a long-term investigation on gastric mucosa by endoscopy. Dig Endoscopy 1992; 4:19–30.
134. Fujioka T, Shuto R, Kodama R, et al. Experimental model for chronic gastritis with *Helicobacter pylori*: long-term follow-up study in *H. Pylori*-infected Japanese macaques. Eur J Gastroenterol Hepatol 1993; 5(Suppl 1):S73–S75.
135. Blaser MJ. *Helicobacter pylori* and the pathogenesis of gastroduodenal inflammation. J Infect Dis 1990; 161: 626–633.
136. Kuipers EJ, Thijs JC, Festen HPM. The prevalence of *Helicobacter pylori* in peptic ulcer disease. Aliment Pharmacol Ther 1995; 9(Suppl 2):59–69.
137. Seppala K. *H pylori* and gastric ulcer disease. In: Hunt RH, Tytgat GNJ, editors. *Helicobacter pylori*. Basic mechanisms to clinical cure. Dordrecht, Netherlands: Kluwer Academic, 1994: 429–436.
138. Graham DY, Lew GM, Klein PD, et al. Effect of treatment of *Helicobacter pylori* infection on the long-term recurrence of gastric or duodenal ulcer: a randomized, controlled study. Ann Intern Med 1992; 116:705–708.
139. Hentschel E, Brandstatter G, Dragasics B, et al. Effect of ranitidine and amoxacillin plus metronidazole in the eradication of *Helicobacter pylori* and the recurrence of duodenal ulcer. N Engl J Med 1993; 328:308–312.
140. Vaira D, Miglioli M, Mule P, et al. Prevalence of peptic ulcer in *Helicobacter pylori* positive blood donors. Gut 1994; 35:309–312.
141. Sipponon P, Varis K, Fraki O, et al. Cumulative 10-year risk of symptomatic duodenal and gastric ulcer in patients with or without chronic gastritis. Scand J Gastroenterol 1990; 25:966–973.
142. Figura N, Guglielmetti P, Rossolini A, et al. Cytotoxin production by *Campylobacter pylori* strains from patients with peptic ulcers and patients with chronic gastritis only. J Clin Microbiol 1989; 27:225–226.
143. Goossens H, Glupczynski Y, Burette A, et al. Role of the vacuolating toxin from *Helicobacter pylori* in the pathogenesis of duodenal and gastric ulcer. Med Microbiol Lett 1992; 1:153–159.
144. Covacci A, Censini S, Bugnoli M, et al. Molecular characterization of the 128-kDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer. Proc Natl Acad Sci USA 1993; 5:695–699.
145. Oderda G, Figura N, Bayeli PF, et al. Serologic IgG recognition of *Helicobacter pylori* cytotoxin associated protein, peptic ulcer and gastroduodenal pathology in childhood. Eur J Gastroenterol Hepatol 1993; 5:695–699.
146. Walsh JH, Peterson WL. The treatment of *Helicobacter pylori* infection in the management of peptic ulcer disease. N Engl J Med 1995; 333:984–991.
147. Martin DF, Montgomery E, Dobeck AS, et al. *Campylobacter*, NSAIDs, and smoking: risk factors for peptic ulcer disease. Am J Gastroenterol 1989; 84:1268–1272.
148. Riccardi VM, Rotter JI. Familial *Helicobacter pylori* infection. Societal factors, human genetics, and bacterial genetics. Ann Intern Med 1994; 120:982–986.
149. Boren T, Falk P, Roth KA, et al. Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. Science 1993; 262:1892–1895.
150. Hook-Nikanen J, Sipponen P, Kosunen TU. Effect of ABO blood group and secretor status on the frequency of *Helicobacter pylori* antibodies. Scand J Gastroenterol 1990; 25:815–818.
151. Azuma T, Konishi J, Tanaka Y, et al. Contribution of HLA-DQA gene to host's response against *Helicobacter pylori*. Lancet 1994; 343:542–543.
152. Calam J. *Helicobacter pylori*, acid, and gastrin. Eur J Gastroenterol Hepatol 1995; 7:310–317.
153. Graham DY, Opekun A, Lew GM, et al. *Helicobacter pylori*- associated exaggerated gastrin release in duodenal ulcer patients. The effect of bombesin infusion and urea ingestion. Gastroenterology 1991; 100:1571–1575.
154. Mulholland G, Ardill JE, Fillmore D, et al. *Helicobacter pylori* related hypergastrinaemia is the result of a selective increase in gastrin 17. Gut 1993; 34:757–761.
155. Moss SF, Legon S, Bishop AE, et al. Effect of *Helicobacter pylori* on gastric somatostatin in duodenal ulcer disease. Lancet 1992; 340:930–933.
156. Graham DY, Lew GM, Lechago J, Antral G-cell and D-cell numbers in *Helicobacter pylori* infection: effect of *H. pylori* eradication. Gastroenterology 1993; 104: 1655–1660.
157. Isenberg JI, Selling HA, Hogan DL, Koss MA. Impaired proximal duodenal mucosal bicarbonate secretion in patients with duodenal ulcer. N Engl J Med 1987; 316: 374–379.
158. Wyatt JI, Rathbone BJ, Dixon MF, Heatley RV. *Campylobacter pyloridis* and acid induced gastric metaplasia in the pathogenesis of duodenitis. J Clin Pathol 1987; 40:841–848.
159. Marshall BJ, Warren JR, Francis GJ, et al. Rapid urease test in the management of *Campylobacter pyloridis*-associated gastritis. Am J Gastroenterol 1987; 82:200–210.
160. Perez-Perez GI, Dworkin BM, Chodos JE, Blaser MJ. *Campylobacter pylori* antibodies in humans. Ann Intern Med 1988; 109:11–17.
161. Graham DY, Klein PD, Evans DJ Jr, et al. *Campylobacter pylori* detected noninvasively by the 13C-urea breath test. Lancet 1987; 1:1174–1177.
162. Marshall BJ, Plankey MW, Hoffman SR, et al. A 20-minute breath test for *Helicobacter pylori*. Am J Gastroenterol 1991; 86:438–445.
163. Graham DY, Evans DG, Evans DJ. Better, accurate, more rapid and convenient physician-office serologic test for detection of *H. pylori* infection. Gastroenterology 1994; 106:A83.
164. Kosunen TU, Seppala K, Sarna S, Sipponen P. Diagnostic value of decreasing IgG, IgA, and IgM antibody titers after eradication of *Helicobacter pylori*. Lancet 1992; 339:.
165. NIH Consensus Development Panel. *Helicobacter pylori* in peptic ulcer disease. J Am Med Assoc 1994; 272:65–69.
166. Buckley M, O'Morain C. Prevalence of *Helicobacter py-*

lori in non-ulcer dyspepsia. Aliment Pharmacol Ther 1995; 9(Suppl 2):53–58.

167. Megraud F. Rationale for the choice of antibiotics for the eradication of *Helicobacter pylori*. Eur J Gastroenterol Hepatol 1995; 7(Suppl 1):S49–S54.
168. Chiba N, Rao BV, Rademaker JW, Hunt RH. Meta-analysis of the efficacy of antibiotic therapy in eradicating *Helicobacter pylori*. Am J Gastroenterol 1992; 87: 1716–1727.
169. Tytgat GNJ. Treatments that impact favorably upon the eradication of *Helicobacter pylori* and ulcer recurrence. Aliment Pharmacol Ther 1994; 8:359–368.
170. Labenz J, Gyenes E, Ruhl GH, Borsch G. Amoxacillin plus omeprazole versus triple therapy for eradication of *Helicobacter pylori* in duodenal ulcer disease: a prospective, randomized, and controlled study. Gut 1993; 34:1167–1170.
171. Bayerdorffer E, Miehlke S, Mannes GA, et al. Double-blind trial of omeprazole and amoxacillin to cure *Helicobacter pylori* infection in patients with duodenal ulcer. Gastroenterology 1995; 108:1412–1417.
172. Laine L, Stein C, Neil G. Limited efficacy of omeprazole-based dual and triple *H. pylori* therapy: a randomized trial employing "optimal" dosing. Gastroenterology 1995; 108:A81.
173. Hosking SW, Ling TK, Yung MY, et al. Randomised, controlled trial of short term treatment to eradicate *Helicobacter pylori* in patients with duodenal ulcer. Br Med J 1992; 305:502–504.
174. Iwahi T, Satoh H, Nakao M, et al. Lansoprazole, a novel benzimidazole proton pump inhibitor, and its related compounds have selective activity against *Helicobacter pylori*. Antimicrob Agents Chemother 1991; 35: 490–496.
175. Graham DY, Lew GM, Klein PD, et al. Effect of treatment of *Helicobacter pylori* infection on the long-term recurrence of gastric or duodenal ulcer. Ann Intern Med 1992; 116:705–708.
176. Chen M, Lee A, Hazell S. Immunisation against gastric Helicobacter infection in a mouse/Helicobacter felis model. Lancet 1992; 339:1120–1121.
177. Chen M, Lee A, Hazell S, et al. Immunisation against gastric infection with Helicobacter species: first step in the prophylaxis against gastric cancer? Int J Med Microbiol Virol Parasitol Infect Dis 1993; 280:155–165.

The Osteoclast and Osteoporosis

Steven L. Teitelbaum, M.D.

Osteoporosis is endemic in western society and ultimately has an impact on one-half of white women attaining 65 years. Osteoporotic fractures typically involve the distal radius, vertebrae, and hip. In fact, Americans sustain approximately 250,000 hip fractures per year and at least 25% of patients with hip fractures die within a year of their episode. Thus, osteoporosis presents itself as a disease of profound and life-threatening consequences.

Regardless of the cause of osteoporosis, the disease always reflects a net loss of skeletal mass and, thus, enhancement of the rate of bone resorption relative to formation. Novel strategies to prevent or arrest osteoporosis depend on understanding the means by which bone is degraded.

The principal, if not exclusive, bone-resorptive cell is the osteoclast, a polykaryon of hematopoietic origin formed by fusion of monocytic precursors. When seen in histological sections, osteoclasts are always juxtaposed to the bone surface and generally found within resorption bays also known as Howship's lacunae.

The most effective means of preventing postmenopausal osteoporosis clearly is estrogen supplementation, which inhibits osteoclastic bone resorption (1). Although it is an effective antiosteoporotic agent, estrogen, enjoys less than optimal long-term patient compliance because of several real or perceived side effects, the most worrisome being enhanced risk of breast cancer. Thus, recent efforts have been particularly directed toward developing nonestrogenic anti-osteoporotic agents.

The fact that the rate of osteoclastic bone resorption dictates development of osteoporosis, particularly in the years immediately after menopause in which bone loss is most rapid (2), and that estrogen exerts its bone-sparing effects by retarding osteoclastic activity (1), prompted exploration of the means by which osteoclasts resorb bone, ultimately aiming to use this information to design skeleton-sparing drugs. With development of techniques whereby osteoclasts can be isolated or generated in culture in sufficient purity and numbers to perform meaningful biochemical and molecular experiments (3, 4), and the identification of several osteopetrotic mutants, the past decade has witnessed major insights into the mechanisms of osteoclastic bone resorption.

The osteopetroses are a family of mutational sclerotic bone diseases reflecting diminished bone resorption caused by failed osteoclastogenesis or generation of dysfunctional osteoclasts (5). Because several distinct molecular defects prompt the osteopetrotic phenotype, a number of rate-limiting phenomena govern the osteoclast's resorptive capacity. Thus, ultimate verification of a particular event is required by the resorptive process and is most definitively obtained when its genetic elimination eventuates in osteopetrosis.

In 1988, Silver et al. (6) observed that the osteoclast forms, at its interface with matrix, an isolated, extracellular microenvironment, which was acidified to a pH approximating 4.5. We established this acidification process as essential to the osteoclast's resorptive capacity by exposing the cell to weak base and, in so doing, arrested bone degradation. Exposed to such acidification, bone mineral is mobilized, thereby exposing bone's organic matrix consisting largely of colla-

---

Presented as a special seminar in the Department of Medicine on May 22, 1995. Accepted May 20, 1996.

Messing Professor of Pathology, Department of Pathology, Washington University School of Medicine, St. Louis, Missouri. Address all correspondence to Steven L. Teitelbaum, M.D., Department of Pathology, Washington University School of Medicine, Barnes-Jewish Hospital, 216 South Kingshighway, St. Louis, MO 63110.

gen fibers (3). Degradation of the skeleton's collagenous component is accomplished, in turn, by osteoclast-secreted enzymes with a highly acidic pH optimum (7). These osteoclast-produced enzymes are lysosomal in nature, a prototype being closely related to cathepsin B. They are delivered to the site of bone resorption by a specific carrier protein known as the mannose 6-phosphate receptor (8). Thus, the osteoclast mediates degradation of both the inorganic and organic phases of bone (3).

The means by which the osteoclast acidifies the resorptive microenvironment is critical to understanding bone degradation. With this in mind we explored the molecular mechanisms by which the osteoclast functions as the most active proton-transporting cell in man. We found acidification of the osteoclast extracellular resorptive microenvironment mediated by an energy-requiring proton pump, similar if not identical to the H^+-ATPase expressed by the intercalated cell of the renal tubule (9, 10). Interestingly, an osteopetrotic-like phenotype exists in man in whom expression of the osteoclast proton pump is deranged, thereby establishing this multidomain complex as an essential component of the cell's resorptive machinery (11). Furthermore, arrest of osteoclast proton pump function prevents bone loss in the oophorectomized rat.

The osteoclast proton transporter resides in the cell membrane juxtaposed to bone (9). It is this polarization of the proton pump into the bone-adjacent plasma membrane which is the *sine qua non* of the resorptive osteoclast. The morphological consequence of proton pump polarization is the ruffled membrane, a complex infolding of plasmalemma formed at the interface of the osteoclast and bone (12). The ruffled membrane, which is the resorptive organelle of the osteoclast, is the product of plasma membrane fusion with acidifying vesicles, probably directionally transported via the microtubular network. Under the aegis of the rab family of small GTPases, these vesicles fuse with the plasma membrane thereby greatly enhancing its complexity (5).

In 1991, Soriano and his colleagues (13) made the surprising observation that deletion of the c-*src* proto-oncogene eventuates in osteopetrosis as its only phenotypic manifestation. Interestingly, osteoclasts are present in normal numbers in the c-*src*-/- mouse but these multinucleated cells are incapable of resorbing bone (14). Ultrastructural examination of the c-*src*-deficient osteoclast reveals that it is incapable of ruffled membrane formation. Thus, c-*src* is essential to the means by which the osteoclast polarizes. Although the precise mechanism by which the proto-oncogene mediates osteoclast polarization is unclear, it appears not to involve tyrosine phosphorylation and may entail facilitated transport of proton pump-bearing vesicles along microtubules. Thus, c-*src* is fundamental to the resorptive process and inhibition of its expression of function presents itself as a potential means of arresting osteoporosis.

The massive extracellular H^+ transport involved in osteoclastic bone resorption necessitates a mechanism whereby base equivalents, in the form of HCO_3^-, derived from the source of intraosteoclastic protons, namely H_2CO_3, are also excreted from the cell to maintain physiological intracellular pH. To address this issue, we once again turned to the paradigm of the renal tubular cell and its energy-independent Cl^-/HCO_3^- exchanger (15). We identified such a cation exchanger on the nonresorptive osteoclast plasma membrane. The Cl^- ions, which enter in exchange for HCO_3^-, move across the cell and exit into the isolated extracellular space, through a ruffled membrane-residing anion channel, charge coupled to the proton pump (16). Thus the osteoclast achieves the remarkably low pH essential for bone degradation by secreting HCl into the resorptive microenvironment (Fig. 1).

The demand for an acidic microenvironment located at the cell-matrix interface, yet isolated from the general extracellular space, necessitates intimacy between the osteoclast and bone. Thus, attention has turned in recent years to the mechanisms by which osteoclasts recognize and attach to skeletal tissue.

Integrins are heterodimeric transmembrane proteins with extracellular matrix binding domains and an intracellular component interacting with the cytoskeleton. Our approach to identifying osteoclast-essential integrins was to distinguish bone matrix proteins to which osteoclasts avidly attach. Osteoclasts bind to osteopontin-1, bone sialoprotein, and vitronectin, all of which bear an arginine-glycine-aspartic acid (RGD) motif recognized by the integrin heterodimer, $\alpha_v\beta_3$ (17). With this information in hand, we asked if the $\alpha_v\beta_3$ integrin is an essential component of the resorptive process. In fact, we found that antibodies to the integrin $\alpha_v\beta_3$ block the capacity of osteoclasts to attach to and degrade bone, respectively, *in vitro*. Thus, expression of the integrin $\alpha_v\beta_3$ by osteoclasts appears to be an essential component to the resorptive process. Most importantly, we have recently identified a nonpeptide mimetic of the $\alpha_v\beta_3$ ligand which obviates bone loss in the oophorectomized rat. Thus,

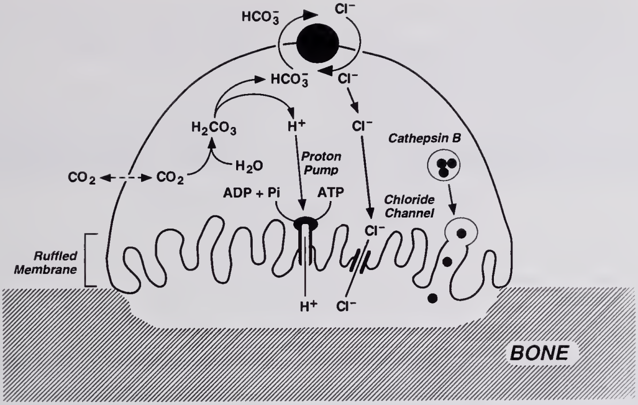


Fig. 1. Osteoclastic bone resorption is initiated by attachment of the cell to bone. Carbon dioxide enters the cell and is hydrated to carbonic acid, which dissociates into bicarbonate ions on protons. To maintain physiological intracellular pH, bicarbonate ions are exchanged for chloride by a non-energy-dependent transporter located at the antiresorptive plasma membrane. The protons, in turn, are formed, via an electrogenic ATPase (proton pump) into an isolated microenvironment located at the interface between the cell and bone. This process diminishes the pH at the site to approximately 4.5. The chloride ions entering the cell are transported into this microenvironment via an anion channel charge coupled to the proton pump. Thus, the cell is transporting hydrochloric acid into the isolated resorptive microenvironment. The low pH obtained at the site mobilizes the mineral phase of bone, thereby exposing organic matrix which, in turn, is cleaved by lysosomal-like enzymes, such as cathepsin B, which are transported via the mannose 6-phosphate receptor to the site of bone degradation.

$\alpha_v\beta_3$ blockade presents itself as a novel approach to prevention of postmenopausal osteoporosis.

Osteoporosis has for centuries remained a common yet enigmatic disease. Although effective prophylaxis in the form of estrogen is in hand, its real and perceived side effects raise demand for alternative approaches. The progress made in bone cell biology during the past decade, particularly the insights gained into the means by which osteoclasts degrade bone, holds promise that the disease, will be preventable by a variety of rational approaches in the foreseeable future.

References

1. Wronski TJ, Cintron M, Doherty AL, Dann LM. Estrogen treatment prevents osteopenia and depresses bone turnover in ovariectomized rats. Endocrinology 1988; 123:681–686.
2. Gallagher JC, Goldgar D, Moy A. Total bone calcium in normal women: effect of age and menopause status. J Bone Miner Res 1987; 2:491–495.
3. Blair HC, Kahn AJ, Crouch EC, et al. Isolated osteoclasts resorb the organic and inorganic components of bone. J Cell Biol 1986; 102:1164–1172.
4. Shioi A, Ross FP, Teitelbaum SL. Enrichment of generated murine osteoclasts. Calcif Tissue Int 1994; 55:387–394.
5. Teitelbaum SL, Abu-Amer Y, Ross FP. Molecular mechanisms of bone resorption. J Cell Biochem 1995; 59:1–10.
6. Silver IA, Murrills RJ, Etherington DJ. Microelectrode studies on the acid microenvironment beneath adherent macrophages and osteoclasts. Exp Cell Res 1988; 175:266–276.
7. Blair HC, Teitelbaum SL, Grosso LE, Lacey DL, Tan H, McCort DW, Jeffrey JJ. Extracellular matrix degradation of acid pH: avian osteoclast acid collagenase isolation and characterization. Biochem J 1993; 290:873–884.
8. Blair HC, Teitelbaum SL, Schimke PA, Konsek JD, Koziol CM, Schlesinger PH. Receptor-mediated uptake of mannose-6-phosphate bearing glycoprotein by isolated chicken osteoclasts. J Cell Physiol 1988; 137:476–482.
9. Blair HC, Teitelbaum SL, Ghiselli R, Gluck S. Osteoclastic bone resorption by a polarized vacuolar proton pump. Science 1989; 245:855–857.
10. Mattsson JP, Schlesinger PH, Keeling DJ, Teitelbaum SL, Stone DK, Xie X. Isolation and reconstitution of a vacuolar-type proton pump of osteoclast membranes. J Biol Chem 1994; 269:24979–24982.
11. Yamamoto T, Kurihara N, Yamaoka K, Ozono K, Okada M, Yamamoto K, Matsumoto S, Michigami T, Ono J, Okada S. Bone marrow-derived osteoclast-like cells from a patient with craniometaphyseal dysplasia lack expression of osteoclast-reactive vacuolar proton pump. J Clin Invest 1993; 91:362–367.
12. Fallon MD, Teitelbaum SL, Kahn AJ. Mutinucleation enhances macrophage-mediated bone resorption. Lab Invest 1983; 49:159–164.
13. Soriano P, Montgomery C, Geske R, Bradley A. Targeted

disruption of the c-src proto-oncogene leads to osteopetrosis in mice. Cell 1991; 64:693–702.

14. Boyce BF, Yoneda T, Lowe C, et al. Requirement of pp60c-src expression for osteoclasts to form ruffled borders and resorb bone in mice. J Clin Invest 1992; 90:1622–1627.
15. Teti A, Blair HC, Teitelbaum SL, et al. Cytoplasmic pH regulation and chloride/bicarbonate exchange in avian osteoclasts. J Clin Invest 1989; 83:227–233.
16. Blair HC, Teitelbaum SL, Tan H, et al. Passive chloride permeability charge coupled to H^+-ATPase of avian osteoclast ruffled membrane. Am J Physiol 1991; 260:C1315–C1324.
17. Ross FP, Alvarez JI, Chappel J, et al. Interactions between the bone matrix proteins osteopontin and bone sialoprotein and the osteoclast integrin $\alpha_v\beta_3$ potentiate bone resorption. J Biol Chem 1993; 268:9901–9907.

Androgen Receptor Mutations in Prostate Cancer

EVELYN R. BARRACK, PH.D.

Role of Androgen Receptors in Prostate Cancer

AMAZINGLY LITTLE IS KNOWN about the prostate other than the serious problem it can create for the aging man. Prostate cancer is now the most commonly diagnosed cancer in men in the United States (1). The prediction for 1995 was 244,000 new cases of prostate cancer; this accounts for 36% of new cancer cases in men (1). Prostate cancer is the second leading cause of deaths from cancer in this group, with 40,000 deaths from prostate cancer in the United States estimated for 1995 (1). The incidence and mortality rates for prostate cancer are higher in African Americans than in Caucasians, for unknown reasons (2). Not only are there racial and ethnic differences in incidence and mortality rates for prostate cancer; there are also dramatic differences in incidence in different countries (2). The highest incidence of prostate cancer is in African Americans, and the lowest incidence is in Japanese and Chinese (2).

Androgen, acting via the androgen receptor (AR), is essential for growth and function of the normal prostate (3, 4). This accounts for my long-standing interest in the role of AR in prostate cancer (5, 6). I see three aspects to this role: carcinogenesis, cancer growth, and progression to androgen independence. Androgen is presumed to play a role in carcinogenesis because males castrated before puberty do not get prostate cancer, and androgen is required in all animal models of prostate carcinogenesis (7, 8). Androgen also plays a role in the growth of prostate cancer because most prostate cancers respond to androgen ablation, the standard treatment for metastatic disease (7). These tumors eventually progress from androgen-dependent to androgen-independent growth (7, 9). Unexpectedly, however, hormone-refractory tumors still express high amounts of AR (10, 11). This was surprising because the prediction had been that cells with AR would be androgen dependent and would die after androgen ablation, leaving AR-negative, androgen-independent cells; it took many years to discover that this premise was incorrect (reviewed in References 5 and 6). A current question of great interest is whether AR might still be active even in this androgen-withdrawn state. Evidence has been published recently that, in the absence of androgen, activation of the AR can be achieved by treating cells with the peptide growth factors insulin-like growth factor-I (IGF-I), keratinocyte growth factor (KGF), or epidermal growth factor (EGF) (12). This phenomenon of steroid-independent receptor activation is not unique to AR. Other steroid receptors also can be activated in the absence of ligand if certain intracellular protein phosphorylation signaling pathways are activated, for example, by 8-bromo-cAMP which activates protein kinase A or by peptide growth factors which activate their tyrosine kinase receptors (13–16).

The question I am addressing is whether these roles of the AR in prostate cancer might be mediated by wild-type AR or by mutant AR.

Androgen Receptor Structure And Function

The AR is a member of the steroid receptor superfamily of ligand-activated nuclear tran-

---

Adapted from the author's seminar presented to the Division of Endocrinology at The Mount Sinai School of Medicine on January 10, 1995.

From the Department of Urology, The Johns Hopkins University School of Medicine, Baltimore, MD. Address reprint requests to Evelyn R. Barrack, Ph.D., Associate Professor of Urology, Department of Urology, The Johns Hopkins University School of Medicine, 600 North Wolfe Street, Marburg Building, Room 115, Baltimore, MD 21287-2101.

This work was supported by grant CA 58236 from the National Cancer Institute.

scription factors that includes receptors for progesterone, glucocorticoid, mineralocorticoid, estrogen, vitamin D, thyroid hormone, and retinoic acid (17, 18). The AR consists of a C-terminal hormone binding domain that confers ligand specificity, a central DNA binding domain that binds to androgen-responsive target genes, and an N-terminal domain that affects transcription efficiency (see Fig. 1) (19, 20). It is remarkable that deletion of the hormone binding domain creates a receptor that is constitutively active, even in the absence of ligand (20). Such a variant, if present in androgen-independent prostate cancer, might account for androgen-independent growth. By analogy, the estrogen receptor (ER) RNA transcript can undergo alternative splicing to yield variant mRNA transcripts that encode ER proteins with altered activity (21–23). For example, an exon 5 deletion variant of the ER encodes a constitutively active ER protein (21, 23). Thus it is conceivable that alternative splicing might also generate a constitutively active AR protein; studies are underway in my laboratory to investigate this possibility. The production of AR mRNA splice variants in prostate cancer has not been studied extensively enough to answer this question (24).

Androgen Receptor Gene Mutations: Androgen Insensitivity Syndrome and LNCaP Cells

Androgen Insensitivity Syndrome. Mutations in the AR gene, which is located on the X-chromosome, were first discovered in patients with androgen insensitivity syndrome (AIS) (25). AIS is an X-linked disease of target organ insensitivity to androgen; it is caused by an AR gene mutation that inactivates receptor function. These are germline mutations, present in every cell in the body, and they are loss-of-function mutations. In complete AIS, individuals with an XY (male) genotype have a female phenotype.

The LNCaP Prostate Cancer Cell Line. The first AR gene mutation found in prostate cancer was in the LNCaP human prostate cancer cell line; amino acid 877 in the hormone binding domain is alanine instead of threonine (26, 27). Unlike AR gene mutations in AIS, the LNCaP mutation does not inactivate AR function. The LNCaP mutant AR binds androgen with an affinity like that of wild-type AR, and, like wild-type AR, is transcriptionally activated by androgen (26, 27). Surprisingly, however, this mutant AR is also transcriptionally activated by estrogen, progesterone, or the antiandrogens cyproterone acetate or hydroxyflutamide (the active metabolite of flutamide) (26, 27). The possibility that this AR gene mutation occurred during establishment of the cell line *in vitro* has not been ruled out. (See Reference 28 for further discussion of this issue.)

Androgen Receptor Mutations in Clinical State B Prostate Cancer: Exons B to H

We were the first to initiate studies to screen for the presence of AR gene mutations in surgical specimens of prostate cancer: clinically localized prostate cancers that were excised by radical prostatectomy (28). Blocks were trimmed to enrich the proportion of tumor in the specimen. By comparing prostate cancer DNA and nontumor DNA, we could determine whether mutations were somatic (present only in the tumor DNA) or germline (present in both). We used polymerase chain reaction (PCR) to amplify genomic DNA and analyzed PCR products for the presence of mutations by using methods that can distinguish mutant AR DNA from wild-type AR DNA (e.g., DNA sequencing, denaturing gradient gel electrophoresis [DGGE], or single-stranded DNA conformation polymorphism analysis [SSCP]).

We reported the first example of a somatic mutation in the AR gene in clinical prostate cancer, a mutation present only in the tumor, not in the patient's lymphocytes (28). The mutation was detected in a radical prostatectomy specimen (Gleason score 8) removed for untreated clinical stage B disease; therefore, it had occurred in the tumor before hormonal therapy. The mutation was a single base change in codon 730 [Val (GTG) → Met (ATG)] (Fig. 1). It was detected both by DGGE and SSCP analysis of PCR products, and its location was defined by sequencing the PCR product (Fig. 2).

Based on the relative intensity of bands in DGGE and SSCP gels, we estimated that the mutation was present in about 50% of the cells in the tumor specimen (28). So it must have occurred in cells with a growth advantage; otherwise it would not have been detectable. If the original cell with the mutation had not had a growth advantage, it would have remained an insignificant percentage of the tumor, becoming diluted by other cells with a growth advantage, and it would not have been detected by screening. The challenging question is whether the AR gene mutation itself conferred a growth advantage on the cell in which it occurred.

Amino acid 730 is located in a region of the hormone binding domain that is highly conserved

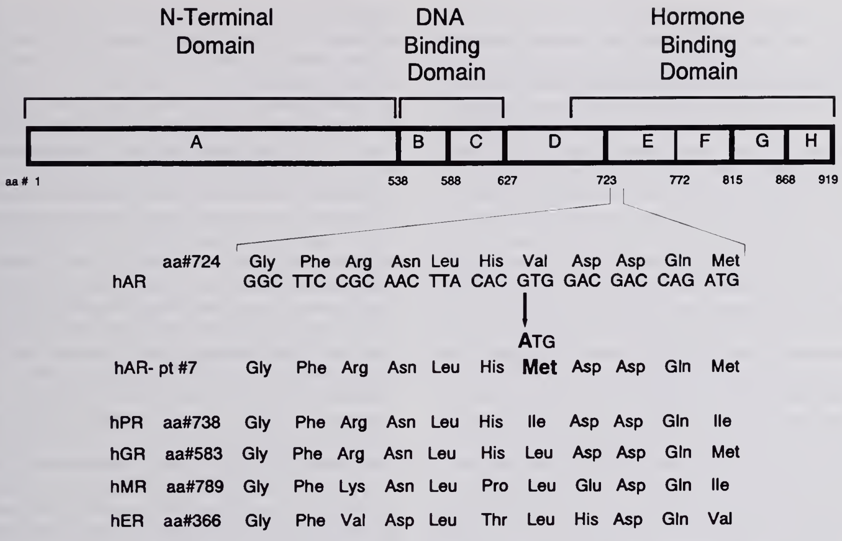

N-Terminal Domain
DNA Binding Domain
Hormone Binding Domain
A
B
C
D
E
F
G
H
aa # 1
538
588
627
723
772
815
868
919
hAR aa#724 Gly Phe Arg Asn Leu His Val Asp Asp Gln Met
GGC TTC CGC AAC TTA CAC GTG GAC GAC CAG ATG
ATG
hAR- pt #7 Gly Phe Arg Asn Leu His Met Asp Asp Gln Met
hPR aa#738 Gly Phe Arg Asn Leu His Ile Asp Asp Gln Ile
hGR aa#583 Gly Phe Arg Asn Leu His Leu Asp Asp Gln Met
hMR aa#789 Gly Phe Lys Asn Leu Pro Leu Glu Asp Gln Ile
hER aa#366 Gly Phe Val Asp Leu Thr Leu His Asp Gln Val

Fig. 1. (Top) Schematic of human AR structure. Amino acid (aa) positions of each exon (labeled A–H) are noted below the bar. Amino acids are numbered according to Reference 17. Codon 877 of a 919-codon AR cDNA (17) is equivalent to codon 868 of a 910-codon AR cDNA (26). The polymorphic glutamine repeat begins at amino acid 58; the glycine repeat begins at amino acid 440–460 (depending on the length of the glutamine repeat). The 8-proline repeat begins at amino acid 363–383 (depending on the length of the glutamine repeat). (Bottom) Comparison of mutant AR (hAR) sequence in stage B prostate cancer (patient 7) (28) with that of wild-type AR and the equivalent region of other steroid hormone receptors: human progesterone receptor (hPR), glucocorticoid receptor (hGR), mineralocorticoid receptor (hMR), and estrogen receptor (hER). Wild-type sequences are from reference 17. Note that the AR gene mutation at codon 730 in patient 7 is in a region of the hormone binding domain highly conserved among steroid receptors.

among different members of the steroid receptor family (Fig. 1). However, codon 730 itself is not conserved in different receptor classes. Interestingly, only the AR contains valine at this codon, and valine is found at this site in human, rat, mouse, and canary AR (17, 29, 30). The presence of a valine at codon 730 uniquely in the AR, and its conservation across species, is a strong indication that valine at this site probably has a critical role in AR function and/or androgen binding, and that the mutation of this amino acid in prostate cancer is likely to affect some aspect of AR function.

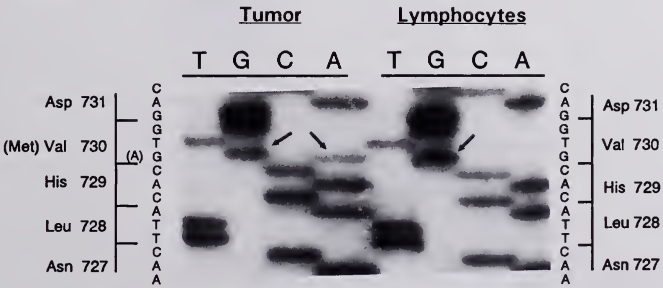

Tumor
Lymphocytes
T G C A
T G C A
Asp 731
(Met) Val 730
(A)
His 729
Leu 728
Asn 727
C A G G T G C A C A T T C A A
C A G G T G C A C A T T C A A
Asp 731
Val 730
His 729
Leu 728
Asn 727

Fig. 2. DNA sequencing gel of the exon E PCR product amplified from the tumor and normal lymphocytes of patient 7. The first nucleotide position of codon 730 is a G, as seen in the patient's peripheral lymphocytes, but the tumor had two bases, G and A, at this position, indicating the presence of a mixture of DNA fragments (wild-type and mutant), diagnostic of a somatic mutation in the tumor (from Reference 28).

Like the AR_{877} mutant in LNCaP cells, the AR_{730} mutant exhibits normal androgen-binding affinity and mediates androgen-dependent transcriptional activation of the MMTV-CAT gene in CV-1 cells (31). Evidence that the transcriptional activity of steroid receptors is highly promoter dependent and cell type specific is extensive (32–34). Therefore, the activity of the AR_{730} mutant on an MMTV viral promoter in monkey kidney CV-1 cells may not be representative of potential effects of this mutant AR in the prostate cancer cells of the patient who harbors the mutation. A mutation at codon 730 has not been found in any patient with complete or partial AIS. This is consistent with our functional characterization, and it suggests that AR gene mutations in prostate cancer [codon 730 (Reference 28) and codon 877 (Reference 26)] may represent a new class of AR mutations that do not cause a loss of function. Our data do not exclude the possibility that the codon 730 mutation may affect the steroid specificity of transactivation, as described for the LNCaP cell line AR mutation at codon 877. A preliminary report suggests that this codon 730 mutation may indeed allow transactivation by a very high concentration of hydroxyflutamide (35). Whether and how this mutation at codon 730 might have affected tumorigenicity in the untreated patient remains a mystery.

What is the frequency of mutations in prostate cancer? In our screen of AR exons B to H in more than 50 clinical stage B prostate cancers, we found several additional mutations (unpublished data), but most of the tumors appear to have no AR gene mutation. This may be caused by the limit of sensitivity of the method used to screen for the presence of mutations; i.e., the smaller the proportion of cells in the specimen with a mutation, the harder it will be to detect the mutation. Alternatively, the mutation frequency may be low in early-stage disease, which is what we have focused on. A low AR gene mutation frequency in early-stage disease would rule out a requisite role of AR gene mutations in the initiation of prostate carcinogenesis, and would indicate that if androgen indeed plays a role in prostate carcinogenesis, then this role is mediated in most cases by a wild-type AR gene. On the other hand, a low AR gene mutation frequency in early-stage disease raises the question of whether there is something unique about those tumors that do have a mutation. One possibility is that these tumors are more aggressive, either because the mutation helped to confer this behavior, or because the mutation occurred in cells with a growth advantage. Of the tumors we have screened to date, the follow-up interval after prostatectomy is too short at this time to be able to say whether the patients with mutations represent a group with more aggressive disease.

Other investigators who have screened for AR mutations in untreated early-stage prostate tumors have found none (24, 36). These reports did not specify whether or not the specimens were enriched for tumor before analysis; the presence of a substantial proportion of nontumor tissue would hamper attempts to detect a mutation. By PCR and SSCP analysis of exons B to H of the genomic DNA, Suzuki et al. (36) found no mutations in seven untreated stage B tumors. Culig et al. (24) used the reverse transcriptase-PCR technique and sequencing to analyze exons B to H of the AR mRNA; they found no mutations in 15 specimens.

Another study reported AR gene mutations in two advanced-stage prostate cancers (one stage C, one stage D); eight other tumors (stage B) had no mutations (37). One mutation was at codon 798 (Gln → Glu); this same mutation has been reported in the germline DNA of a patient with partial AIS (38). All other AR mutations detected in prostate cancer are unique to prostate cancer, and have not been reported to occur in patients with androgen insensitivity syndrome.

Exon A of the Androgen Receptor

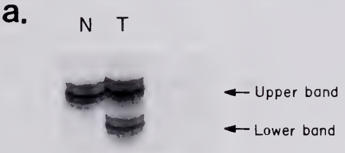
Because exon A encodes about 60% of the full-length AR, we recognized the need to study it before drawing conclusions about mutation frequency. Exon A encodes the N-terminal domain, the part of the molecule that affects transcriptional efficiency. A distinguishing feature of the AR protein is that the N-terminal domain contains repeats of glutamines, prolines, and glycines (17, 18) (see Fig. 1). Glutamine repeats, glutamine-rich domains, and proline repeats are an important feature of several transcription factors (39, 40). For example, the glutamine-rich domain of the transcription factor Sp1 is critical for activity, and it mediates interaction with another essential component of the transcription apparatus, namely the TATA box-binding protein (40). Glutamine repeats can form β-sheets held together strongly by hydrogen bonds, so they might function as "polar zippers" and mediate protein-protein interactions (41).

A fascinating feature of the glutamine and glycine repeats in the AR is that their lengths vary in the population (i.e., are polymorphic) (42–45). These polymorphisms are encoded by CAG repeats (range, 11–31) and GGC repeats (range, 10–22), respectively. Therefore, AR with different CAG (glutamine) and GGC (glycine) repeat lengths represent different AR alleles or AR variants, all apparently normal because they are found in the general population. The proline repeat appears to be constant in length.

The length of the glutamine repeat affects the transcriptional activity of the AR. Deletion of the glutamine repeat yields an AR with increased transcriptional activity, and an increase in glutamine repeat length decreases activity (46, 47). The role of the glycine repeat and proline repeat in AR function is unknown. Another fascinating aspect of this CAG repeat is that it is twice the normal length in the germline DNA of patients with X-linked spinal and bulbar muscular atrophy (SBMA, also known as Kennedy's disease), an adult-onset neurological disorder (43). Androgen action is somewhat blunted in men with this AR defect (40–52 CAGs), because the ability of a synthetic androgen to lower plasma testosterone, luteinizing hormone, and follicle-stimulating hormone levels is significantly lower in SBMA patients than in controls (48). In addition, an AR with a glutamine repeat that is twice the normal length (as found in patients with SBMA) has half as much activity in stimulating androgen-responsive gene transcription (47).

The AR CAG repeat is also interesting because microsatellite DNA sequences are potential sites of genetic instability (49). In addition to SBMA, several other neurological diseases (fragile-X syndrome, myotonic dystrophy, Huntington's disease, spinocerebellar ataxia type 1, and dentatorubral-pallidoluysian atrophy) have also been linked to microsatellite expansions in the genes that are thought to be involved in disease development (50). In addition, widespread microsatellite instability caused by defects in mismatch repair enzymes occurs in many colon cancers and other types of cancers (49).

We therefore screened prostate cancers for evidence of expansion or deletion of the CAG microsatellite in the AR gene. We amplified the part of the gene that encodes the repeat, and analyzed PCR product length on a sizing gel (51). Because the length of the CAG repeat varies within the population (42, 44), detection of a somatic change in repeat length in a tumor requires concurrent evaluation of nontumor tissue from the same patient. The AR gene is on the X-chromosome (25), so males have only a single allele per cell, which accounts for the presence of a single major PCR product from each specimen (Fig. 3a, lane N). In one patient, nontumor DNA contained a single allele, as expected, but his tumor DNA contained two alleles, one identical to normal and a second smaller one (Fig. 3a, lane T). We cut out the bands, sequenced them, and discovered that the single allele in the nontumor tissue contained 24 CAGs (Fig. 3b). The tumor contained one allele with 24 CAGs (wild-type) and a second with 18 CAGs (mutant) (Fig. 3b). So the tumor contains cells with a somatic, in-frame contraction of the AR CAG repeat ($CAG_{24} \rightarrow CAG_{18}$). We wonder whether the CAG_{18} mutation might have enhanced AR function in the patient's prostate can-



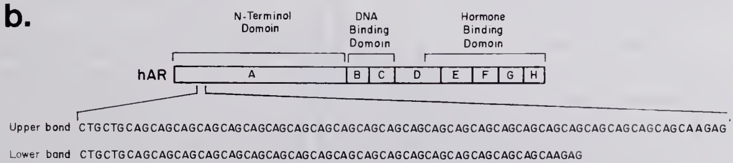

b.
N-Terminal Domain
DNA Binding Domain
Hormone Binding Domain
hAR
A
B
C
D
E
F
G
H
Upper band CTGCTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAAGAG
Lower band CTGCTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAAGAG

Fig. 3. (a) Prostate cancer with a somatic mutation in the CAG repeat of the AR gene. Nontumor DNA yielded a single major PCR product (upper band), whereas tumor DNA yielded two major products (upper and lower band), indicating the presence of cells with a wild-type allele (upper band) and cells with a mutant allele (lower band) that contains a deletion. Accompanying minor bands are a common feature in denaturing gels of PCR products of repeat sequences. (b) A schematic representation of the AR notes the location and sequence of the CAG repeat in the upper band (wild-type, present in tumor and nontumor DNA) and lower band (mutant, present only in tumor). The upper band product contains 24 CAGs; the lower band product contains 18 CAG repeats (Adapted from Reference 51.)

cer cells, because AR with shorter glutamine repeats are more active (46, 47).

The patient with the $CAG_{24} \rightarrow CAG_{18}$ mutation has an interesting clinical history (51). The mutation was present in his radical prostatectomy specimen (Gleason score 8), which was removed for untreated clinical stage B disease. Radical prostatectomy was followed by radiation therapy because the surgical margins contained tumor. Three years later the patient's serum prostate-specific antigen (PSA) level rose to 190 ng/mL, and a bone scan revealed evidence of metastatic disease. After treatment with an LHRH agonist and the antiandrogen flutamide, his PSA level fell to castrate levels (<0.2 ng/mL). Eighteen months later his PSA level rose to 35 ng/mL, despite a serum testosterone level still in the castrate range, apparently signaling tumor progression to androgen independence. Unexpectedly, however, cessation of flutamide therapy led to a decline in serum PSA to 7.5 ng/mL within 1 month, which suggests that flutamide was acting paradoxically as an agonist; this phenomenon has also been described by others (52, 53). The basis for the paradoxical response to flutamide remains to be determined, as well as whether glutamine repeat length could affect binding and transcriptional activity by hydroxyflutamide. In addition, other mutations in the AR gene may have occurred during therapy; we are unable to obtain tumor tissue from this patient because he is in remission, so we are unable to analyze his tumor AR in its current state.

Because this mutation is a microsatellite mutation, we wondered whether it might be a marker of widespread genetic instability. We think this is unlikely, however, because microsatellite instability appears to be uncommon in sporadic prostate cancer (54). In addition, of more than 40 tumors screened, only one had an AR CAG repeat length mutation (51).

Androgen Receptor Mutations in Hormone Refractory Prostate Cancer

AR mutations have been found in the hormone binding domain in several hormone refractory prostate cancers (36, 37, 55–57). Culig et al. (55) used reverse transcriptase-PCR and sequencing to screen seven aspiration biopsies of metastatic prostate cancer; they found a mutation in one hormone refractory tumor. The mutation was at codon 715 (Val → Met); it did not affect the steroid specificity of ligand binding and it did not disrupt androgen-dependent transcriptional activity, but it did increase the efficacy (maximal effect) with which progesterone and adrenal steroids like dehydroepiandrosterone and androstenedione could activate transcriptional activity of the AR (55). Suzuki et al. (36) performed autopsies on eight patients with hormone refractory prostate cancer and found an AR gene mutation in one. The primary tumor contained a codon 701 mutation (Leu → His), properties unknown, whereas a codon 877 mutation, identical with the mutation found in the LNCaP cell line, was found in the liver, kidney, and lymph node metastases of the same patient (36). Codon 877 may be a hot spot for mutations, having been found in 6 of 24 microdissected hormone-refractory tumors (56). Of course, the possibility that PCR products amplified from prostate specimens might become contaminated in the laboratory with LNCaP mutant AR PCR product is recognized and must be rigorously avoided. Taplin et al. (57) found AR mutations in the RNA of metastatic deposits of four of eight patients with hormone refractory prostate cancer. One of the mutations was at codon 877, but it was a Thr → Ser mutation (57), different from the Thr → Ala mutation found in LNCaP cells (26); both mutations at this codon broaden the steroid specificity of the AR, enhancing the extent of transcriptional activation by R5020 (a progestin) or estradiol (58). A mutation at codon 874 (His → Tyr) also enhanced the extent of transcriptional activation of the AR by estradiol (57). It is not known whether AR gene mutations in hormone-refractory disease arose during androgen ablation therapy or were already present prior to therapy, like those we found in untreated stage B tumors. It will be interesting to determine whether AR mutations found in early-stage and late-stage disease affect AR function in similar ways.

Androgen Receptor Mutations in Latent Prostate Cancer

Based on autopsy studies, about 15 to 30% of men older than 45 to 50 years of age have latent prostate cancer, i.e., histological evidence of prostate cancer (59). The prevalence of latent prostate cancer is the same in Japanese men living in Japan and American men living in the United States (59), but the incidence of clinical prostate cancer is many times lower in Japanese men than in American men (2). In a recent preliminary study, 16 of 74 (22%) Japanese men with latent prostate cancer were reported to have somatic AR gene mutations (60). Eight of these mutations would ablate AR activity either because they create a premature stop codon or because the same codon has been found to be mutated in the germline DNA of unrelated patients with AIS. The effect of the other mutations on AR function was not characterized. Maybe the occurrence of a somatic AR gene mutation that interferes with androgen action precludes progression to clinical disease. AR mutations were not found in Americans with latent prostate cancer (60); maybe this accounts for the higher rate of progression to clinical prostate cancer in American men. This preliminary report by Takahashi et al. (60) deserves follow-up.

Proposal: Polymorphic Variants Of the AR Modulate Androgen Action in Prostate Cancer

Although studies are still underway to determine the overall frequency of AR gene mutations in prostate cancer, the presence of AR gene mutations in only a subset of early-stage prostate cancers suggests that the putative role of androgen in prostate carcinogenesis is mediated in most cases by a wild-type AR gene, which exists as multiple alleles that differ in glutamine and glycine repeat length.

It is interesting that the frequency in the general population of AR alleles with different CAG repeat lengths varies in different racial groups that have a different incidence of prostate cancer (42) (see Fig. 4). In Caucasians, the most common AR allele has 21 CAGs, whereas in African Americans, who have a significantly higher prostate cancer incidence and mortality rate (2), the most common allele has 18 CAGs (42) (Fig. 4). The frequency distributions actually appear to be bimodal (Fig. 4).

Because an abnormal length of the glutamine repeat affects transcriptional activity, AR with longer glutamine repeats having lower activity (46–48), we propose that the normal polymorphism of the AR glutamine repeat length in the general population may also affect AR activity. Androgen plays an important role in development, growth, and behavior, so normal AR polymorphism may help to account for variations in the degree of androgenization and manifestations of androgenic activity in the general population. Some androgen-mediated processes that might be affected are aggression, hair growth, and penile length. Potential differences in the AR activity of "normal" AR alleles could even affect the role of androgen in prostate cancer. Thus, the existence of multiple allelic variants, with a range of activity that may depend on the glutamine and/or glycine repeat length, may help to account for the fact that racial groups with different AR allele frequency distributions (42) have different incidence rates of prostate cancer (2). AR with shorter glutamine repeats may have more activity, and the effect of androgen in prostate cancer may be greater in individuals with such AR alleles. If short repeats enhance AR activity and if AR activity is required for prostate carcinogenesis, then maybe the higher frequency of shorter AR CAG repeats (e.g., CAG_{18}) in African Americans (Fig. 4) accounts for the higher frequency of prostate cancer in African American men. Studies cur-

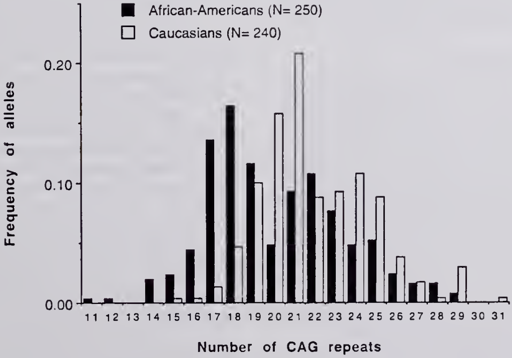


Fig. 4. Frequency distributions of AR CAG repeat length in the general population of African Americans and Caucasians (based on data from Reference 42).

rently underway in my laboratory suggest that the frequency of AR alleles with a short CAG repeat is higher in men with prostate cancer than in the general population (61).

Summary

AR gene mutations occur in both early-stage prostate cancers (28, 51, our unpublished data) and late-stage disease (36, 37, 55–57). One common feature is that both types of mutations retain ligand-dependent transcriptional activity. We speculate that AR mutations may characterize a more aggressive disease, or confer an ability to survive androgen ablation therapy. A large percentage of tumors appears to have no AR gene mutation, but the possibility has not been ruled out that tumors without an AR gene mutation may nonetheless produce variant AR, for example, by alternative splicing. The apparent absence of AR gene mutations in the majority of early-stage tumors indicates that the role of androgen in the development of clinical prostate cancer is mediated predominantly by a normal AR gene, which exists as multiple alleles that differ in glutamine and glycine repeat length, and that potentially differ in signal-transducing activity. Glutamine and glycine repeat length may thereby modulate the effect of androgen on tumor cell proliferation. The effect of glutamine and glycine repeat length on AR function may determine the sensitivity of tumor cells to existing tissue levels of dihydrotestosterone, but tissue dihydrotestosterone levels depend on circulating androgen levels and the amount of 5α-reductase activity in the tissue. Therefore, although potential AR activity may be affected by the length of the glutamine and/or glycine repeat, actual AR activity will depend also on these other factors.

Acknowledgments

I acknowledge the people in my laboratory (past and present) who have actually done the work I described: Drs. Jay Newmark, Diane Hardy, Mark Schoenberg, Rachel Rondonelli, and Janette Hakimi. I also acknowledge the contributions of my colleagues: Drs. Jonathan Epstein, Bill Isaacs, and Steve Bova who provided prostate cancer DNA samples, Patrick Walsh who brought to our attention patients with an interesting clinical history, and Terry Brown who taught us about DGGE and helped us to characterize the transcriptional activity of mutant AR. I thank all of these people.

References

1. Wingo PA, Tong T, Bolden S. Cancer statistics, 1995. CA Cancer J Clin 1995; 45:8–30.
2. Ross RK, Paganini-Hill A, Henderson BE. The etiology of prostate cancer: What does the epidemiology suggest? Prostate 1983; 4:333–344.
3. Cunha GR, Donjacour AA, Cooke PS, et al. The endocrinology and developmental biology of the prostate. Endocrine Rev 1987; 8:338–362.
4. Luke MC, Coffey DS. The male sex accessory tissues. Structure, androgen action, and physiology. In: Knobil E, Neill JD, editors. The physiology of reproduction, 2nd ed. New York: Raven Press, 1994: 1435–1487.
5. Barrack ER, Tindall DJ. A critical evaluation of the use of androgen receptor assays to predict the androgen responsiveness of prostatic cancer. Prog Clin Biol Res 1987; 239:155–187.
6. Sadi MV, Barrack ER. Androgen receptors in prostate cancer. pp. 239–265 In: Pretlow II TG, Pretlow TP, editors. Biochemical and molecular aspects of selected cancers, vol. 2. New York: Academic Press, 1994: 239–265.
7. Catalona WJ, Scott WW. Carcinoma of the prostate. In: Walsh PC, Gittes RF, Perlmutter AD, Stamey TA, editors. Campbell's Urology, 5th ed. Philadelphia: WB Saunders, 1986: 1463–1534.
8. Wilding G. The importance of steroid hormones in prostate cancer. Cancer Surv 1992; 14:113–130.
9. Scott WW, Menon M, Walsh PC. Hormonal therapy of prostatic cancer. Cancer 1980; 45:1929–1936.
10. van der Kwast TH, Schalken J, Ruizeveld de Winter JA, et al. Androgen receptors in endocrine-therapy-resistant human prostate cancer. Int J Cancer 1991; 48:189–193.
11. Ruizeveld de Winter JA, Janssen JA, Sleddens HMEB, et al. Androgen receptor status in localized and locally progressive hormone refractory human prostate cancer. Am J Pathol 1994; 144:735–746.
12. Culig Z, Hobisch A, Cronauer MV, et al. Androgen receptor activation in prostatic tumor cell lines by insulin-like growth factor-I, keratinocyte growth factor, and epidermal growth factor. Cancer Res 1994; 54:5474–5478.
13. Denner LA, Weigel NL, Maxwell BL, et al. Regulation of progesterone receptor-mediated transcription by phosphorylation. Science 1990; 250:1740–1743.
14. Power RF, Mani SK, Codina J, et al. Dopaminergic and ligand-independent activation of steroid hormone receptors. Science 1991; 254:1636–1639.
15. Aronica SM, Katzenellenbogen BS. Stimulation of estrogen receptor-mediated transcription and alteration in the phosphorylation state of the rat uterine estrogen receptor by estrogen, cyclic adenosine monophosphate, and insulin-like growth factor-I. Mol Endocrinol 1993; 7:743–752.
16. Ignar-Trowbridge DM, Teng CT, Ross KA, et al. Peptide growth factors elicit estrogen receptor-dependent transcriptional activation of an estrogen-responsive element. Mol Endocrinol 1993; 7:992–998.
17. Lubahn DB, Joseph DR, Sar M, et al. The human androgen receptor: complementary deoxyribonucleic acid cloning, sequence analysis and gene expression in prostate. Mol Endocrinol 1988; 2:1265–1275.
18. Chang C, Kokontis J, Liao S. Structural analysis of cDNA and amino acid sequences of human and rat androgen receptors. Proc Natl Acad Sci USA 1988; 85:7211–7215.
19. Simental JA, Sar M, Lane MV, et al. Transcriptional activation and nuclear targeting signals of the human androgen receptor. J Biol Chem 1991; 266:510–518.
20. Jenster G, van der Korput HAGM, van Vroonhoven C, et al. Domains of the human androgen receptor involved in steroid binding, transcriptional activation, and subcellular localization. Mol Endocrinol 1991; 5:1396–1404.
21. Fuqua SAW, Fitzgerald SD, Chamness GC, et al. Variant

human breast tumor estrogen receptor with constitutive transcriptional activity. Cancer Res 1991; 51:105–109.

22. Wang Y, Miksicek RJ. Identification of a dominant negative form of the human estrogen receptor. Mol Endocrinol 1991; 5:1707–1715.
23. Castles CG, Fuqua SAW, Klotz DM, Hill SM. Expression of a constitutively active estrogen receptor variant in the estrogen receptor-negative BT-20 human breast cancer cell line. Cancer Res 1993; 53:5934–5939.
24. Culig Z, Klocker H, Eberle J, et al. DNA sequence of the androgen receptor in prostatic tumor cell lines and tissue specimens assessed by means of the polymerase chain reaction. Prostate 1993; 22:11–22.
25. Quigley CA, De Bellis A, Marschke KB, et al. Androgen receptor defects: Historical, clinical, and molecular perspectives. Endocrine Rev 1995; 16:272–321.
26. Veldscholte J, Ris-Stalpers C, Kuiper GGJM, et al. A mutation in the ligand binding domain of the androgen receptor of human LNCaP cells affects steroid binding characteristics and response to anti-androgens. Biochem Biophys Res Commun 1990; 173:534–540.
27. Veldscholte J, Berrevoets CA, Brinkmann AO, et al. Anti-androgens and the mutated androgen receptor of LNCaP cells: differential effects on binding affinity, heat shock protein interactions, and transcriptional activation. Biochemistry 1992; 31:2393–2399.
28. Newmark JR, Hardy DO, Tonb DC, et al. Androgen receptor gene mutations in prostate cancer. Proc Natl Acad Sci USA 1992; 89:6319–6323.
29. Charest NJ, Zhou Z-X, Lubahn DB, et al. A frameshift mutation destabilizes androgen receptor messenger RNA in the Tfm mouse. Mol Endocrinol 1991; 5:573–581.
30. Nastiuk KL, Clayton DF. Seasonal and tissue-specific regulation of canary androgen receptor messenger ribonucleic acid. Endocrinology 1994; 134:640–649.
31. Schoenberg MP, Tonb DC, Brown TR, Barrack ER. Functional analysis of an androgen receptor gene mutation in localized human prostate cancer. Abstract presented at the 7th Annual Meeting of the Society for Basic Urologic Research; May 14–15, 1993; San Antonio, Texas.
32. Vegeto E, Shahbaz MM, Wen DX, et al. Human progesterone receptor A form is a cell- and promoter-specific repressor of human progesterone receptor B function. Mol Endocrinol 1993; 7:1244–1255.
33. Ince BA, Montano MM, Katzenellenbogen BS. Activation of transcriptionally inactive human estrogen receptors by cyclic adenosine 3′,5′-monophosphate and ligands including antiestrogens. Mol Endocrinol 1994; 8:1397–1406.
34. Tzukerman MT, Esty A, Santiso-Mere D, et al. Human estrogen receptor transactivational capacity is determined by both cellular and promoter context and mediated by two functionally distinct intramolecular regions. Mol Endocrinol 1994; 8:21–30.
35. Stober J, Culig Z, Gast A, et al. Ligand specific modulation of activity of mutant androgen receptors from human prostate cancer: a model for investigating hormone-receptor interactions. Abstract presented at AACR Special Conference on Basic and Clinical Aspects of Prostate Cancer; December 8–12, 1994; Palm Springs, CA.
36. Suzuki H, Sato N, Watabe Y, et al. Androgen receptor gene mutations in human prostate cancer. J Steroid Biochem Mol Biol 1993; 46:759–765.
38. Batch JA, Williams DM, Davies HR, et al. Androgen receptor gene mutations identified by SSCP in fourteen subjects with androgen insensitivity syndrome. Hum Mol Genet 1992; 1:497–503.
39. Gerber H-P, Seipel K, Georgiev O, et al. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. Science 1994; 263:808–811.
40. Emili A, Greenblatt J, Ingles CJ. Species-specific interaction of the glutamine-rich activation domains of Sp1 with the TATA box-binding protein. Mol Cell Biol 1994; 14:1582–1593.
41. Perutz MF, Johnson T, Suzuki M, Finch JT. Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. Proc Natl Acad Sci USA 1994; 91:5355–5358.
42. Edwards A, Hammond HA, Jin L, et al. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 1992; 12:241–253.
43. La Spada AR, Wilson EM, Lubahn DB, et al. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 1991; 352:77–79.
44. Macke JP, Hu N, Hu S, et al. Sequence variation in the androgen receptor gene is not a common determinant of male sexual orientation. Am J Hum Genet 1993; 53:844–852.
45. Sleddens HFBM, Oostra BA, Brinkmann AO, Trapman J. Trinucleotide (GGN) repeat polymorphism in the human androgen receptor (AR) gene. Hum Mol Genet 1993; 2:493.
46. Chamberlain NL, Driver ED, Miesfeld RL. The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. Nucleic Acids Res 1994; 22:3181–3186.
47. Mhatre AN, Trifiro MA, Kaufman M, et al. Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. Nature Genet 1993; 5:184–188.
48. Sobue G, Doyu M, Morishima T, et al. Aberrant androgen action and increased size of tandem CAG repeat in androgen receptor gene in X-linked recessive bulbospinal neuronopathy. J Neurol Sci 1994; 121:167–171.
49. Loeb LA. Microsatellite instability: marker of a mutator phenotype in cancer. Cancer Res 1994; 54:5059–5063.
50. Martin JB. Molecular genetics of neurological diseases. Science 1993; 262:674–676.
51. Schoenberg MP, Hakimi JM, Wang S, et al. Microsatellite mutation ($\text{CAG}_{24\rightarrow18}$) in the androgen receptor gene in human prostate cancer. Biochem Biophys Res Commun 1994; 198:74–80.
52. Scher HI, Kelly WK. Flutamide withdrawal syndrome: its impact on clinical trials in hormone-refractory prostate cancer. J Clin Oncol 1993; 11:1566–1572.
53. Sartor O, Cooper M, Weinberger M, et al. Surprising activity of flutamide withdrawal, when combined with aminoglutethimide, in treatment of “hormone refractory” prostate cancer. J Natl Cancer Inst 1994; 86:222–227.
54. Bussemakers MJG, Bova GS, Schoenberg MP, et al. Microsatellite instability in human prostate cancer (abstract 967). J Urol 1994; 151:469A.
55. Culig Z, Hobisch A, Cronauer MV, et al. Mutant androgen receptor detected in an advanced-stage prostatic carcinoma is activated by adrenal androgens and progesterone. Mol Endocrinol 1993; 7:1541–1550.
56. Gaddipati JP, McLeod DG, Heidenberg HB, et al. Frequent detection of codon 877 mutation in the androgen

receptor gene in advanced prostate cancer. Cancer Res 1994; 54:2861–2864.

57. Taplin M-E, Bubley G, Balk S. Molecular analysis of the androgen receptor in androgen independent prostate cancer. Presented at Basic and Clinical Aspects of Prostate Cancer, An AACR Special Conference; December 8–12, 1994; Palm Springs, CA.

58. Ris-Stalpers C, Verleun-Mooijman MCT, Trapman J, Brinkmann AO. Threonine on amino acid position 868 in the human androgen receptor is essential for androgen binding specificity and functional activity. Biochem Biophys Res Commun 1993; 196:173–180.

59. Carter HB, Piantadosi S, Isaacs JT. Clinical evidence for and implications of the multistep development of prostate cancer. J Urol 1990; 143:742–746.

60. Takahashi H, Furusato M, Allsbrook WC, Boyd J. Inactivating mutations of the androgen receptor gene in latent prostatic carcinomas (abstract 1621). Proc Am Assoc Cancer Res 1994; 35:271.

61. Hakimi JM, Schoenberg MP, Barrack ER. Androgen receptor variants as a potential risk factor for aggressive prostate cancer (abstract 216). J Urol 1995; 153:282A.

Childhood Immunizations—1996

David S. Hodes, M.D.

The most dramatic moments in the development of childhood vaccines occurred earlier in this century when death rates from diphtheria, pertussis, measles, and polio underwent spectacular decreases. Nonetheless, the current scene is still marked by the steady introduction of new vaccines against other prominent childhood diseases and the refinement of existing vaccines. These current developments are reviewed briefly.

Hepatitis B. Although the impact of this disease is not as great in the United States as in many Asian and African countries, it is still estimated that 300,000 cases occur here each year (1). Discovery of the causative agent of hepatitis B led to the systematic production of an effective vaccine. The current vaccine is not a whole virus, but a "subunit" vaccine consisting of a single viral protein, in this case the surface antigen (HBsAg). Antibody raised against this antigen provides protection against all subtypes of the virus. Earlier versions of the vaccine consisted of DNA-free, noninfectious HBsAg purified from the sera of infected patients. Formalin treatment added a further safety step. Although these versions were effective, they have been replaced by preparations in which the HBsAg has been synthesized in yeast containing plasmids bearing the HBsAg gene.

Despite the fact that effective vaccines became available in 1982 and were recommended for individuals at high risk, little impact was made on the incidence of hepatitis B in this country, possibly because of increased immigration from endemic areas. Consequently, at the end of 1991, the United States Public Health Service recommended universal vaccination of all infants. Current practice calls for an initial vaccination at birth before discharge from the hospital, a second immunization between 1 and 2 months, and a third between 6 and 18 months. An alternative schedule is a series of immunizations at 1 to 2, 4, and 6 to 18 months (2). This latter schedule cannot be applied to the child of the mother with known infection. Here the infant must be vaccinated at birth and the vaccination supplemented with the administration of hepatitis B hyperimmune globulin.

Varicella. The development of live attenuated varicella zoster virus (VZV) vaccines is not new. The most commonly administered vaccine was produced in Japan in the 1970s (3) and has been used for a long time there. Resistance to the general use of VZV vaccines in the United States stemmed from a handful of concerns and perceptions. First, like other members of the herpes virus family, VZV produces a lifelong persistent infection and, when activated, results in a painful, occasionally debilitating disease, zoster. There was concern that attenuated VZV vaccine virus strains might be more prone to produce zoster than do wild-type strains. Second, VZV in the immunocompromised host is occasionally a serious, life-threatening infection. There was concern that the inadvertant spread of vaccine virus to immunocompromised hosts might pose a substantial risk. Finally, there was the general belief that chickenpox, the condition produced by primary infection with VZV, is a trivial illness.

It has taken the better part of 20 years for these roadblocks to be overcome. As best as can be determined, the incidence of zoster after the injection of VZV vaccine is substantially less than that seen after infection with wild-type virus (4), no case of clinical disease has ever been caused by the spread of vaccine virus from inoculated

---

Professor of Pediatrics, Division of Pediatric Infectious Diseases, Mount Sinai School of Medicine.

Address requests for reprints to Dr. Hodes at Mount Sinai School of Medicine, Box 1198, One Gustave L. Levy Place, New York, NY 10029.

healthy volunteers (5), and epidemiologic studies remind us that infection with VZV is not always harmless. Complications of chickenpox including varying degrees of mild scarring at the site of lesions, bacterial superinfections, Reye syndrome, pneumonia, encephalitis (most frequently presenting as cerebellar ataxia), glomerulonephritis, and arthritis (6). Approximately 90 fatal cases are reported annually to the Centers for Disease Control (CDC) with most of the fatalities occurring in the pediatric age group (7). It is estimated that the annual national health cost of chickenpox is $399 million, most of it in lost parent workdays (8).

The currently marketed live attenuated VZV vaccine is a cell-free preparation of the Japanese OKA strain and is produced in the United States by Merck. It is attenuated by tissue culture passage, mostly in human fibroblasts. Universal vaccination of children aged 12 to 18 months with a single dose of subcutaneously administered vaccine is now recommended (8). In addition, children 18 months to 13 years of age without previous vaccination and lacking a reliable history of chickenpox should be similarly vaccinated. Children older than 13 years of age with similar history are given two doses 4 to 8 weeks apart. Adolescents older than 18 years of age with a negative history for chickenpox may, if desired, be tested for immunity by an enzyme immunoassay or latex agglutination assay and the vaccination foregone if the test is positive.

Haemophilus influenzae **type b.** *H. influenzae* type b (Hib) was, until recently, one of the most prevalent and feared bacteria in pediatrics. It is the cause of epiglottitis and a frequent cause of pneumonia, septic arthritis, and meningitis as well. Hib has exhibited a characteristic epidemiologic pattern regarding age groups. Newborns, presumably because of the effect of maternally derived antibody, are protected, but children up to the age of 5 years exhibit marked susceptibility. Naturally acquired immunity seems to protect older children. Disease caused by Hib was not uncommon before the present era, with invasive disease by age 5 years having a cumulative risk of 1 in 200. This included an annual national meningitis caseload of 12,000 (9).

The successful development of the current Hib vaccines is a landmark in several aspects. It marks the first occasion in which a vaccine designed for universal administration has been effective in the prevention of a bacterial disease produced by direct bacterial invasion of the bloodstream and internal organs. It also is the first such vaccine to achieve enhanced immunogenicity by biochemical manipulation of the antigen, in this case the capsular polysaccharide, against which protective antibody is mounted.

Initial Hib vaccines were developed after it was recognized that the major virulence factor and antigen against which protective antibody is directed is the capsular polysaccharide, polyribosylribitol phosphate (PRP). Pure preparations of PRP, however, proved nonantigenic in children younger than 18 months of age, thus leaving the children most likely to acquire invasive disease unprotected. This problem was overcome by developers in the late 1980s when PRP was linked chemically to certain protein carriers, thereby enhancing its immunogenicity in young infants and making possible universal immunization programs (10). Currently, three preparation are licensed for use in the United States. In one preparation, a large-chain PRP is conjugated to the outer membrane protein of *Neisseria meningitidis;* in a second, a medium chain PRP is conjugated to diphtheria toxoid; and, in a third, PRP oligosaccharide is conjugated to a mutant nontoxic diphtheria toxin. There are variations in vaccination schedule according to the vaccine preparation, but Hib vaccination is given contemporaneously with vaccinations for diphtheria, pertussis, tetanus, and polio, and the first series of inoculations is completed within the first 6 months of life.

The results of universal Hib vaccination generally have been dramatic. British studies, in intention-to-treat analysis, have estimated an efficacy of 90% protection against disease (11), whereas data on 114,000 infants studied in Finland indicated 94% protection (12), and data from the CDC document indicated an average annual decline in Hib mortality of 48% in each of the 3 years after the introduction of the conjugated vaccines (13). The most recent data indicate that the rate of invasive Hib disease nationwide in children younger than 5 years of age declined from more than 40 cases per 100,000 population in 1987 to 1.7 per 100,000 in 1994 (14). Oddly, the vaccine has been less effective in Eskimo children, a group particularly susceptible to Hib disease (15).

Hepatitis A. Poor reporting limits our knowledge of the exact extent of hepatitis A infection in the United States (16), but the approximately 25,000 cases reported annually to the CDC (17) could easily represent an underestimate of an order of magnitude or more.

The elucidation of the cause of hepatitis A, like that of hepatitis B, is a relatively recent development. As in hepatitis B, scientific discoveries were followed by a methodical expansion of new knowledge culminating in the production of an apparently effective vaccine. Hepatitis A virus was first identified in 1973 by electron microscopic examination of stools from patients with the disease (18). The agent is a small, nonenveloped virus containing a genome of single-stranded RNA and is classified with the enteroviruses, although some of its physical and genetic properties are somewhat atypical for the group (19).

Vaccine development became feasible when it became possible to cultivate the virus in tissue culture. Recently, a vaccine produced by Smithkline Beecham was licensed for use in the United States. This is an inactivated virus preparation, produced by growing the virus in tissue culture, effecting purification through ultrafiltration and chromatography, and using formalin as the inactivating agent. A product produced by Merck is similar.

Field trials of inactivated hepatitis A virus vaccines in children, so far, have proved the vaccines to be both immunogenic and free of significant side effects (20–22). Clinical efficacy has also been impressive (23, 24). Werzberger et al. (23) sought to control an outbreak of hepatitis A in a Hasidic community in upstate New York by means of a single injection of inactivated vaccine. Within 3 weeks of immunization no further cases of hepatitis were encountered among 519 children who received the vaccine, whereas 34 cases developed during the next 4 months among the 518 children who received placebo. Antibody concentrations achieved exceeded those attainable with the passive administration of immunoglobulin and were persistent as well. Worldwide experience with inactivated hepatitis A vaccines now includes more than 50,000 children and adults with vaccine regimens proving virtually 100% immunogenic (25).

Despite the optimistic assessments thus far, neither the American Academy of Pediatrics nor the Advisory Committee on Immunization Practices of the CDC has yet recommended universal immunization. Perhaps this step will be taken when more experience has been garnered with the combined hepatitis A-hepatitis B vaccines currently under development. At the moment, hepatitis A vaccine is approved by the Food and Drug Administration for use in individuals older than 2 years of age, especially for travelers to endemic areas. The vaccinee is given an intramuscular injection and is considered protected after 4 weeks. Permanent immunity is provided by a second injection 6 to 12 months later (26). If universal vaccination of children is adopted, the vaccine will be administered according to the same schedule currently used for hepatitis B.

Pertussis. The pertussis vaccine traditionally used in the United States is a "whole-cell" vaccine consisting of an inactivated suspension of *Bordatella pertussis,* combined with diptheria and tetanus toxoid and adsorbed onto an aluminum salt. The vaccine has received much unjustified notoriety as a result of the widespread confusion of temporal association and causality. Abroad, this notoriety led to the temporary discontinuance of universal vaccination in Britain and Japan during the late 1970s. The result of these moves was a resurgence of the disease in both countries with more than 100,000 cases and at least 41 deaths (27). Although most of the now-past controversy about the whole-cell vaccine centered on the alleged severe and permanent neurologic damage supposedly caused by the preparation, it is clearly not an optimal vaccine. First, the protection rate is hardly impressive, being estimated variously from 40 to 90% (28); and, second, the occurrence of annoying minor side effects, such as fever, resultant febrile convulsions, inflammation at the injection site, and irritability is common (27). A self-limited hypotonic-hyporesponsive state is also reported on occasion.

The attribution of severe and permanent encephalopathy to pertussis immunization stems from the 1948 report of Byers and Moll who reported neurologic disorders in 15 children shortly after the administration of pertussis vaccine (29). In 1958, Berg's review noted 107 case reports in the literature in which pertussis vaccine was associated with neurologic reactions (30), and in 1974 Kulenkampff and colleagues (31) reported 36 children with neurologic reactions after administration of the combinationdiphtheria-pertussis-tetanus (DPT) vaccine. Reading these early reports should rouse the suspicions of the skeptical. The conditions reported are clinically diverse with only seizures being a fairly constant feature, and there are virtually no neuropathologic studies to confirm that one is dealing with a specific entity. A series of surveillance studies followed that variously estimated the incidence of permanent neurologic damage from the pertussis vaccine at between 1 per 50,000 and 1 per 400,000 vaccinated children (27).

More sophisticated epidemiologic studies in Britain, however, raised some questions. A 7-year longitudinal cohort study involving 134,700 children receiving DPT vaccine and 133,500 receiving only a diptheria-tetanus combination, revealed four instances of neurologic disease occurring in the former group and none in the latter. Of the four cases, however, all could be attributed to concommitant acute viral infections. The famous National Childhood Encephalopathy Study (NCES) had a case-control format and involved all British children with acute neurologic disorders seen over a 3-year period (32). The NCES produced the oft-quoted claim that the incidence of serious neurologic disease in otherwise healthy children is 1 per 330,000 pertussis immunizations. More careful scrutiny, however, indicated problems with this interpretation (33, 34). Misclassification resulting in incorrect assignment of a number of cases to small cells converted insignificant associations to significant ones. With the correction of these cell sizes, it was impossible to demonstrate that pertussis vaccination was a risk factor for encephalopathy. The case against whole-cell pertussis vaccine as a cause of permanent brain damage is now regarded as unproven and conditions regarded as contraindications to vaccination now downgraded to indications for precautions (35).

Current pertussis vaccine development, therefore, should now be regarded as an attempt to improve protection from disease and minimize discomfort rather than as a campaign to prevent mental retardation as previously viewed in the lay press. Clear progress is being made with current acellular preparations. These are, rather than preparations of inactivated whole bacteria, purifications of proteins deemed significant in the pathogenesis and/or immunogenicity of *B. pertussis*. Vaccines of single proteins and combinations have been manufactured and tested, and their efficacy has generally been similar. These vaccines have all included an inactivated form of the pertussis toxin (also known as the lymphocytosis-promoting factor) and other antigens included have been the filamentous hemagglutinin, agglutinogens, and an outer-membrane protein known as periactin. Two acellular vaccines are currently licensed in the United States, one containing filamentous hemagglutinin and inactivated pertussis toxin, and the other containing all four antigens. Because the acellular preparations have shown their greatest antigenicity in older children, current recommendations call for infants to be vaccinated with whole-cell vaccines for the initial 2, 4, and 6-month inoculations with acellular vaccines an option for the 12- to 18-month and 4- to 6-year boosters (36). The goal of reducing side effects with the acellular preparations seems to have been attained; and, using the recommended schedule, antigenicity is equivalent to that seen using only whole-cell preparations. Long-term studies on the equivalency of protection against actual disease are not yet available.

Polio. Few chapters in the history of medicine read as dramatically as that relating the conquest of polio, so it is somewhat surprising that there is interest in revising the current spectacularly successful regimen. A measure of this success is seen in CDC records which show a total of 21,269 cases of paralytic poliomyelitis in 1952 and three in 1993 (37).

The customary polio vaccine given in the United States is the trivalent, live-attenuated vaccine administered orally (OPV). It is given at 2 and 4 months of age with boosters recommended any time between 6 and 18 months and at 4 to 6 years. OPV replaced the inactivated Salk preparation (IPV) between 1961 and 1963. Authorities at that time noted OPV had several advantages. It avoids injection, making it more acceptable to parents and children; it induces local intestinal immunity; and the OPV strain itself is shed for extended periods of time in the stool of vaccinees. This last property allows for vaccinees to spread the attenuated strains to unvaccinated children. This is no small advantage in dealing with underprivileged populations where delivery of medical care is often incomplete. It doubtless played an important role in the Pan American Health Organization's successful campaign to wipe out polio in Latin America using OPV. With widespread use of OPV the circulation of wild-type virus ceases.

The performance record of OPV in the United States is evident from the data quoted. Furthermore, it seems to have been an advance over IPV. Salk vaccine had reduced the incidence of polio by about 90% between its introduction in 1955 and the early 1960s (37). This was an admirable accomplishment but was less than the virtually total elimination achieved after the switch to OPV. It has been argued that the eradication of poliomyelitis with IPV is feasible only in small, economically developed and ethnically homogeneous countries like those of northern Europe. A further drawback of the IPV was the potential for industrial accidents. As the Salk vaccine involved wild-type (WT) poliovirus, failure in the inactivation process could lead to the inoculation of children with live pathogenic virus. This occurred in the disastrous Cutter incident in 1955 with resultant paralytic polio in 94 vaccinees and 166 of their contacts (38). Finally, in countries relying on IPV there have been epidemics when WT virus was introduced into the population (39, 40).

Still, there is also an imperfection in the OPV program. Because of partial reversion to virulence of the attenuated virus strain, paralytic poliomyelitis will rarely develop in vaccinees or their contacts. The risk is greater after the first administration of vaccine (about 1/560,000 doses) than after subsequent doses (1/10.5 million) (42). This incidence means that about 8 to 10 cases of vaccine-attributable paralysis (VAP) annually in the United States. Most frequently affected are adult close contacts who constitute two thirds of the cases (43) and the immunodeficient. The problem raised by these cases has brought about a careful reconsideration of American immunization practices, and changes may be expected.

The case for inactivated vaccines has been bolstered by several factors. First, improved industrial techniques have rendered repetitions of the Cutter incident highly unlikely. Second, the development in the 1980s of IPVs of enhanced antigenic potency (EIPVs) has led to the marketing of vaccines more effective than the earlier products that sported the imperfect records. There is now a well-documented experience in shifts from OPV to EIPV programs.

France shifted from OPV to EIPV in 1983, spurred by the intention of eliminating VAP. France, like the United States, is a large country with a heterogeneous population. Further it is subjected to large numbers of visitors from parts of the Eastern hemisphere where, unlike the Western hemisphere, WT poliovirus still circulates. France has not seen an endogenous case of paralytic polio since 1989. Thus the French population has been protected against spread of WT virus from immigrants. Further, surveillance of sewerage in France has failed to reveal WT virus (44). It is tempting to conclude with certainty that EIPV has succeeded in protecting France from imported disease and in keeping WT poliovirus from circulation. It must be remembered, however, that only a fraction of Frenchmen attained their immunity from polio by means of EIPV. Whether EIPV alone can create a protective herd immunity in the face of introduction of WT virus is less clear. IPV has only an incomplete effect in blocking fecal excretion of WT virus (43), but this may be sufficient to prevent epidemics in developed countries where hygiene is at a high level. Some doubts remains, but the record to date is encouraging.

A final development that begins to shift the balance toward EIPV is the elimination of WT virus from general circulation throughout the Western hemisphere. This has doubtless lessened the level of herd immunity needed to protect against disease and softened the requirement for the high levels of herd immunity that require an OPV program to effect. When there is little or no virus in a community, less general immunity is needed to protect against it. The question facing the United States now is whether eliminating the risk that EIPV is incapable of protecting the country against WT virus imported from the Eastern hemisphere is worth the price of 40 to 50 VAPs. This is the number of VAPs that will occur before the year 2000, the target date the World Health Organization has set for the global elimination WT poliovirus by using OPV in the Eastern hemisphere.

At the moment, change appears to be imminent. At its latest meeting, the Advisory Committee on Immunization Practices of the Centers for Disease Control recommended that the 2- and 4-month vaccinations be carried out with EIPV, with OPV being reserved for the 6- to 18-month and 4- to 6-year boosters (45). This would eliminate the vast majority of VAPs, which are usually associated with first vaccinations, and would, at the same time, preserve the herd immunity advantage of OPV by inducing intestinal immunity. The opinion of the American Academy of Pediatrics is expected to follow suit.

Streptococcus pneumoniae. Interest in vaccination against this pathogen in children has been lying dormant for a long time. Now, because of the emergence of penicillin-resistant strains of *S. pneumoniae* (pneumococcus), there are signs of increased activity.

Immunologic immaturity in infants and young children presents a problem for the development of pneumococcal vaccines. The importance of type-specific antibody to the capsular polysaccharide of the pneumoncoccus has long been recognized. This antibody is thought to interact with the complement system to facilitate phagocytosis of the invading bacteria. Unfortunately, children, younger than 2 years of age respond poorly to polysaccharide antigens; diminished responses to previously produced pneumococcal polysaccharide vaccines have been noted in children up to the age of 5 (45, 46). Unfortunately, this youngest age group is the one in greatest need of protection from invasive pneumococcal disease such as meningitis, bacteremia, and pneumonia.

One obstacle to the development of a successful antipneumococcal vaccine is the superfluity of capsular serotypes of *S. pneumoniae* and the apparent type specificity of antibody-mediated protection. This represents a difficulty not faced in the case of Hib. The problem becomes more manageable, however, when it is recognized that, despite the fact that there are at least 83 pneumococcal serotypes, from 60 to 90% of invasive disease is produced by just a few of them (47, 48). Thus vaccines directed at the most clinically significant strains could prevent the bulk of pneumococcal disease even though they will not be able to eliminate it entirely. A theoretical caveat is that such a vaccination program, based on the strains implicated by current epidemiology, could lead to the selection of other strains which might be just as troublesome in the end.

The problem of poor infant response to polysaccharide antigens is being tackled by the pharmaceutical industry via strategies derived from the Hib program, namely, coupling of bacterial polysaccharides to protein carriers to enhance immunogenicity. One such vaccine from Lederle-Praxis Biologicals combines the polysaccharides of five common pathogenic serotypes (6B, 14, 18C, 19F, and 23F) with mutant diptheria toxin (CRM197). Trials are planned for children particularly susceptible to pneumococcal infection such as those infected with human immunodeficiency virus.

References

1. Centers for Disease Control. Protection against viral hepatitis: Recommendations of the Immunization Practices Advisory Committee (ACIP) MMWR 1990; 39(no. RR-2):1–23.
2. Centers for Disease Control. A comprehensive strategy for the elimination of hepatitis B virus transmission in the United States through universal childhood immunization: Recommendations of the Immunization Practices Advisory Committee (ACIP). MMWR 1991; 40(no. RR-13):1–25.
3. Takahashi M, Otsuka T, Okuno Y, et al. Live vaccine used to prevent the spread of varicella in children in hospital. Lancet 1974; 2:288–290.
4. Hardy IB, Gershon A, Steinberg S, et al. The incidence of zoster after immunization with live attenuated varicella vaccine. A study in children with leukemia. N Engl J Med 1991; 325:1545–1550.
5. Weibel RE, Neff BJ, Kuter BJ, et al. Live attenuated varicella vaccine. Efficacy trial in healthy children. N Engl J Med 1984; 310:1409–1415.
6. Fleisher GW, Henry W, McSorley M, et al. Life-threatening complications of varicella. Am J Dis Child 1981; 135:896–899.
7. Committee on Infectious Diseases, American Academy of Pediatrics. Recommendations for the use of live attenuated varicella vaccine. Pediatrics 1995; 95:791–796.
8. Preblud SR, Orenstein WA, Koplan JP, et al. A benefit-cost analysis of a childhood vaccination program. Postgrad Med J 1985; 61:17–22.
9. Cochi SL, Broome CV, Hightower MS. Immunization of US children with *Haemophilus influenzae* type b polysaccharide vaccine. J Am Med Assoc 1985; 253:521–529.
10. Robbins JB, Schneerson G. Polysaccharide-protein conjugates: a new generation of vaccines. J Infect Dis 1990; 161:821–832.
11. Booy R, Hodgson S, Carpenter L. Efficacy of *Haemophilus influenzae* type b conjugate vaccine PRP-T. Lancet 1994; 344:362–366.
12. Eskola J, Käyhty H, Takala AD, et al. A randomized, prospective field trial of a conjugate vaccine in the protection of infants and young children against invasive *Haemophilus influenzae* type b disease. N Engl J Med 1990; 323:1381–1387.
13. Schoendorf KC, Adams WG, Kiely JL, et al. National trends in *Haemophilus influenzae* meningitis mortality and hospitalization among children, 1980 through 1991. Pediatrics 1994; 93:663–668.
14. Centers for Disease Control. Progress toward elimination of *Haemophilus influenzae* type b disease among infants and children—United States, 1993–1994. MMWR 1995; 44:545–550.
15. Ward J, Brenneman G, Letson GW, et al. Limited efficacy of a *Haemophilus influenzae* type b conjugate vaccine in Alaska native infants. N Engl J Med 1990; 323:1393–1401.
15. Melnick JL. History and epidemiology of hepatitis A virus. J Infect Dis 1995; 171(Suppl 1):S2–S8.
17. Centers for Disease Control. Summary of notifiable diseases, United States 1993. MMWR 1994;42:1–73.
18. Feinstone SM, Kapikian AZ, Purcell RH. Hepatitis A: detection by immune electron microscopy of a virus-like antigen associated with acute illness. Science 1973; 182:1026–1028.
19. Gust ID, Coulepis AG, Feinstone SM, et al. Taxonomic classification of hepatitis A virus. Intervirology 1983; 20:1–7.
20. Horng Y-C, Chang M-H, Lee C-Y, et al. Safety and immunogenicity of hepatitis A vaccine in healthy children. Pediatr Infect Dis J 1993; 12:359–362.
21. Block SL, Hendrick JA, Tyler RD, et al. Safety, tolerability, and immunogenicity of a formalin-inactivated hepatitis A vaccine (VAQTA) in rural Kentucky children. Pediatr Infect Dis J 1993; 12:976–980.
22. Balcarek KB, Bagley MR, Pass RF, et al. Safety and immunogenicity of an inactivated hepatitis A vaccine in preschool children. J Infect Dis 1995; 171:(Suppl 1):S70–S72.
23. Werzberger A, Mensch B, Kuter B, et al. A controlled trial of a formalin-inactivated hepatitis A vaccine in healthy children. N Engl J Med 1992; 327:453–457.
24. Innis BL, Snitbhan R, Kunasol P, et al. Protection against hepatitis A by an inactivated vaccine. J Am Med Assoc 1994; 271:1328–1334.
25. Clemens G, Safara A, Hepburn A, et al. Clinical experience with an inactivated hepatitis A vaccine. J Infect Dis 1995; 171(Suppl 1):S44–S49.
26. Centers for Disease Control. Licensure of an inactivated hepatitis A vaccine and recommendations for use among international travellers. MMWR 1995; 44:559–560.
27. Cherry JD, Brunell PA, Golden GS, et al. Report of the task force on pertussis and pertussis immunization—1988. Pediatrics 1988; 81(Suppl):939–984.
28. Fine PEM, Clarkson JA. Reflections on the efficacy of pertussis vaccines. Rev Infect Dis 1987; 9:866–883.
29. Byers RK, Moll FC. Encephalopathies following prophylactic pertussis vaccine. Pediatrics 1948; 1:437–457.
30. Berg JM. Neurologic complications of pertussis immunization. Br Med J 1958; 2:24–27.
31. Kulenkampff M, Schwartzman JS, Wilson J. Neurological

complications of pertussis inoculation. Arch Dis Child 1974; 49:46–49.

32. National Childhood Encephalopathy Study. Whooping Cough: Reports from the Committee on Safety of Medicines and the Joint Committee on Vaccination and Immunization. London; Department of Health and Social Security, Her Majesty's Stationery Office, 1981.
33. Bowie C. Lessons from the pertussis vaccine court trial. Lancet 1990; 335:397–399.
34. Cherry JD. "Pertussis vaccine encephalopathy": It is time to recognize it as the myth that it is. JAMA 1990; 263: 1679–1680.
35. Committee on Infectious diseases, American Academy of Pediatrics. 1994 Red Book: Report of the Committee on Infectious Diseases. Elk Grove Village, IL: American Academy of Pediatrics, 1994, 355–367.
36. Centers for Disease Control. Summary of notifiable diseases, United States 1994. MMWR 1995; 43:69–80.
37. Salk D. Eradication of poliomyelitis in the United States. II Experience with killed poliovirus vaccine. Rev Infect Dis 1980; 2:243–257.
38. Nathanson N, Langmuir AD. The Cutter incident: poliomyelitis following formaldehyde-inactivated poliovirus vaccination in the United States during the spring of 1955. Am J Hyg 1963; 78:16–81.
39. Hovi T, Cantell K, Jhouvilainen A, et al. Outbreak of paralytic poliomyelitis in Finland: Widespread circulation of antigenically altered poliovirus type 3 in a vaccinated population. Lancet 1986; 1:1427–1432.
40. Slater PE, Orenstein WA, Morag A, et al. Poliomyelitis outbreak in Israel in 1988: A report with two commentaries. Lancet 1990; 335:1192–1198.
41. Nkowane BM, Wassilak SGF, Orenstein WA, et al. Vaccine-associated paralytic poliomyelitis: United States, 1973–1984. JAMA 1987; 257:1335–1340.
42. Moore M, Katona P, Kaplan JE, et al. Poliomyelitis in the United States, 1969–1981. J Infect Dis 1982; 146:558–563.
43. Plotkin SA. Inactivated polio vaccine for the United States: a missed vaccination opportunity. Pediatr Infect Dis J 1995; 14:835–839.
44. The New York Times, Oct. 19, 1995.
45. Douglas RM, Paton JC, Duncan SJ, et al. Antibody response to pneumococcal vaccination in children younger than five years of age. J Infect Dis 1983; 148: 131–137.
46. Paton JC, Toogood IF, Cockington RA, et al. Antibody response to pneumococcal vaccine in children aged 5 to 15 years. Am J Dis Child 1986; 140:135–138.
47. Austrian R, Howie VM, Ploussard JH. The bacteriology of pneumococcal otitis media. Johns Hopkins Med J 1977; 141:104–111.
48. Orange M, Gray B. Pneumococcal serotypes causing disease in children in Alabama. Pediatr Infect Dis J 1993; 12:244–245.

Physicians' Reactions to Patients:

What Has Happened during the Past 10 Years

RICHARD GORLIN, M.D., JAMES STRAIN, M.D., AND ROSAMOND RHODES, PH.D.

Abstract

New stresses have been added to the medical life of physicians in training during the past decade. These include the newly diverse and complex sociocultural environment of the patient, new and renascent lethal diseases, widespread substance abuse, and the effects of the breakdown of the fabric of our society. These factors complicate the practice of medicine as never before. Educators must keep pace and prepare young physicians to understand and then cope with these major new forces.

IN 1981 GORLIN AND ZUCKER produced a film, "A Complicating Factor: Doctors' Feelings As a Factor in Medical Care" (1), to help medical students sort through their personal responses to patients and to understand what patients may be communicating indirectly.

In 1983 Gorlin and Zucker further described the reaction of the physician to the patient and how this affects both the physician and the quality of the relationship of the physician with the patient (2). At that time the major issues concerned the ability of the physician to reckon with his or her own reactions to patients who presented with problems or character traits that flawed the relationship. Such patients included those who were fundamentally hateful and unlikable, who denied their disease state or who became unusually dependent upon the physician, and sometimes those patients who were intimidating to the doctor. The article by Gorlin and Zucker considered some of the behavioral mechanisms by which impulses harmful to the physician-patient relationship might be controlled. In the process of achieving emotional awareness and mastery, the physician's own sense of comfort and self-esteem with his/her role could be maintained.

Since that time major sociocultural forces have arisen in society which affect the physician in ways not even imagined in 1980. Poverty, substance abuse, new cultures, religions, and social mores have become more common among our patient populations. This is further exacerbated by the appearance of new and old lethal diseases. Both teachers and students are remarkably unprepared for coping with the patient of the 1990s.

"Cultural Indigestion"

The past decade has seen unprecedented immigration to this country. Simple "digestion" or incorporation of the large numbers of persons who have immigrated, predominantly to about seven states, who need to be brought into the mainstream of American life, has been a problem. Legal immigration between 1971 and 1990 was about 10.5 million people, with an additional conservatively estimated 3 million illegal immigrants. Immigration in 1991 included nearly 2

Dr. George Baehr Professor of Clinical Medicine, Senior Vice President (RG); Professor of Psychiatry, Director, Division of Behavioral Medicine, and Consultation Psychiatry (JS); and Associate Professor of Medical Education, Director of Bioethics Education (RR), The Mount Sinai Medical Center, New York, New York.

Address reprint requests to Richard Gorlin, M.D., Senior Vice President, The Mount Sinai Medical Center, Box 1018, One Gustave L. Levy Place, New York, 10029.

million legal immigrants and an additional 500,000 estimated illegal immigrants (3). Assimilation has been complicated by the fact that the flood of immigration has come from the widest possible diversity of cultures, races, and religious beliefs that has ever descended upon any one nation during so short a time. On a typical day, the Elmhurst Hospital Center, an affiliate of the Mount Sinai School of Medicine, may field requests for interpreters of more than 20 languages for patients who cannot speak English and who derive from the multiple populations settled in the Borough of Queens, New York. Not only has the patient population changed, but the house staff has changed as well. The percentage of European-descended, American-born physicians is shrinking in relation to those from other backgrounds (Table).

The receiving areas for new immigrants have, by the very nature of immigration, been limited so that there has been an unprecedented clustering of races, ethnic groups, and cultures. The main targets have been large urban areas in those states close to immigrating population bases or which have served as traditional gateways to this country [California, Arizona, Texas, Illinois, New York, Florida, New Jersey (3, 4)]. Although these population centers carry the heaviest burden in accommodating new arrivals, the John Deere phenomenon is also still at work. Wherever there are jobs, particularly for unskilled labor, newcomers will find them. Many towns away from the big centers now have sizable racially and ethnically diverse populations.

The situation is further complicated by the fact that probably one quarter of all immigrants are not legally admitted to this country, which decidedly alters their attitudes toward contact with the government and with health care facilities. All too many arrive without the skills needed for getting along in a modern, technologically oriented society. Illegal aliens can not only be defensive and hostile but may become isolated within their own cultural group. Furthermore, our society is sometimes unprepared to help them for such deceptively simple reasons as lack of adequate numbers of interpreters and the perception that the new arrivals are "different."

TABLE

Patient and Housestaff Ethnic Breakdown at Mount Sinai Affiliates (New York)

	Patient Population (%)	Housestaff Population (%)
Mount Sinai		
European/American	56	71
Other	44	29
Elmhurst		
European/American	12	19
Other	78	81
Queens Hospital Center		
European/American	10	14
Other	90	84

In prior generations immigrating groups often rushed to learn English and to adopt American customs and ways, even though they gathered in specific areas of a city. For a variety of reasons, some populous immigrating groups have elected to maintain separate identities rather than to assimilate or integrate. Thus, there has been a perpetuation of the use of the native language without the acquisition of English as well. Such singular attachment to the cultural base can virtually shut such individuals off from the economic and social mainstream of the country.

This is in stark contrast to the many ethnic and racial groups which came to this country between 1880 and 1920, seeking assimilation and working, not with government assistance, which was unavailable in that era, but through local, independent social agencies organized by each racial or ethnic group. A persistent example to this day is the manner in which the Korean community maintains its own social and financial support system for new immigrants. During the 1880 to 1920 immigration many newcomers moved on to seek work far beyond the gateway cities in the steel mills, coal mines, and farm lands. The final feature about the wave of immigration around the turn of the century was that this was not a particularly diverse group. These people were for the most part European and Caucasian and followed the broad elements of a Judaeo-Christian philosophy.

This changed in the 1980s. We are now a nation of the most remarkable diversity ever seen. This offers a unique opportunity for the country by bringing talents from an even wider racial, social, and cultural base. Things have happened so quickly, however, that isolation, intergroup collision, and overuse of limited resources have created a disequilibrium.

For a doctor to know English and a little Spanish is no longer adequate. No longer does it suffice to understand the basic precepts of a few religions. Now Islam, Buddhism, animism, and other religions influence the behavior and responses of the patients.

Potentially Lethal Diseases, New and Old

Physicians during this past decade have encountered totally new and devastating diseases, such as human immunodeficiency virus (HIV). Parasitic diseases from Asia, Africa, and some parts of Latin American have been reintroduced or have newly surfaced in the United States. Drug-resistant strains of tuberculosis, a disease that had been virtually wiped out, are a serious public health scourge. The impact of these infectious diseases on physicians who had looked upon the era of contagion as long past in North America has been overwhelming at times.

Societal Disorders

The medical aspects of HIV and tuberculosis are covered in the medical school curriculum but the intrinsic societal aspects are virtually ignored. Young physicians must also cope with the adverse medical impact of substance abuse. This has extended beyond alcohol and heroin to include "crack cocaine." These substances have been used by people in all socioeconomic strata and ethnic groups, with enormous negative impact on behavior, and potentiation of lethal diseases, such as HIV and hepatitis. The sheer number of patients so afflicted is enormous. The need is compelling for the medical resident to understand and to cope not only with the diseases but also with issues, such as complex or varying forms of sexuality, drug dependency, and styles of "family" as well as of individual life. Socially negative behavior is further exacerbated when patients are homeless, beset with mental illness, and lack family, religious, and social support.

Cultural, Religious, and Attitudinal Impediments to Traditional Medical Care

A patient at Mount Sinai, in whom tuberculosis had been diagnosed, refused to wear his mask because in his culture, a man never covered his face, particularly his mouth. He also refused to take his antituberculosis medication as prescribed. When remanded by the court to 3 months in an acute-care hospital bed, the medical staff were furious that a noncompliant patient would consume essential medical resources and subject them to contagion as well. They felt that his belief system prevented him from deriving benefit from his medical care and that the situation was entirely unfair.

Encountering many patients who do not speak English, who may not have been in a Western-style hospital, clinic or office, and who may not know what to expect of an American physician, places a major stress on the young physician and hence on the patient-physician relationship itself.

Does the physician know how the patient from another society or culture feels about our principles of practice? Our country is rife with alternative belief systems for healing. Most doctors and hospitals follow the allopathic (or osteopathic) principles which have dominated American medicine for more than a century. Consequently, there are few other "legitimate" options available, yet requests for care in other venues are increasing. Not only Jehovah's Witnesses and Christian Scientists but those (especially from nonwestern cultures) who have little belief or trust in allopathic medicine come to us for care. Frequently they then reject our offerings, and yet will not leave our premises. Patients from third-world societies or from areas where education in medical matters is relatively unsophisticated present a special problem to the physician: Seemingly routine proposals for treatment may not be understood, and hence not accepted. A person from another culture or belief system has difficulty enough adjusting to a new world without having to cope with some of the medical options we offer, even when there is no countervailing philosophy to challenge them. Are we educating physicians to recognize these beliefs and attitudes, not so much that they be accepted but that they be understood?

Patient Autonomy versus Accountability

Our society respects patients' autonomy in making decisions about their bodies and their health. Presumably, such decisions are based on adequate information and the patient's competence to form an opinion. Conflict often occurs when a physician proposes a treatment that does not conform to a different culture or belief system. Patients may respond with distrust as well as disagreement with our proposals for management.

Our house officers were stymied by a patient who would not allow the administration of intravenous fluids because she said it was an unnatural way for anything to enter the body. She wanted instead to burn incense in her room and have the medicines spread on her food. Despite the physician's protestations that oral medication would not be effective, the patient insisted that the medication be spread on cold chicken wings that her family had dried in the sun for the last 48 hours.

Does such a patient have an obligation to be more trusting, compliant, or at least accountable, to the physician who tries to help? Distrust and hostility can frustrate (or even frighten) the physician, who can in turn be impatient and hostile and feel hopelessly ineffectual in using his or her considerable medical knowledge. Can we better educate the patient about the health care system, expectations he might have, and who the players are? It would be useful for health care planners to deal with the responsibilities of the patient as well as those of the doctor, nurse, insurer, and government. Many of the medical problems we see in daily practice are rooted in pervasive social problems. Teen pregnancy, drunk driving, drug abuse, violence, and family dysfunction are not immediately rectifiable by health care legislation. As U.S. Senator Daniel Patrick Moynihan has stated, “[T]rying to fix the health care delivery system directly might not be worth the trouble if nothing [is] done to cure the social ills that increase medical costs” (5).

The Patient-Physician Encounter

No matter what the medical, social, and cultural factors are, no matter what the massive economic plans put forth to achieve health coverage in this nation, it all boils down to a room containing the patient and the doctor. This should be a positive interaction with the physician approaching the patient with sympathy, understanding, and confident authority. For the reasons given above, this is often not the case. There are thunderclouds of unspoken distrust, lack of comprehension, and irrational miscommunication hovering over the relationship and raining down on the attempt to treat illness. Furthermore, we often encounter the patient only in an episode of illness and not in a continuum of care. We see the patient at the nadir of the physical plight, and can intervene in only a small portion of the patient’s overall problems, most of which transcend the medical system. Yet we are asked to assume this unrealistic burden. Physicians often are held responsible for “fixing” society. Some analysts even suggest, “holding providers accountable for improving the health status of their communities” (6). This twist of logic would hold doctors responsible for disease and for the setting in which it occurs! The White House and the Congress should look at how to alter the sociocultural context that breeds and magnifies illness even as they attempt to redesign the health care system.

The Physician’s Dilemma

The physician is neither prepared nor authorized to deal with problems arising from economic, sociological, and cultural roots. Unfamiliar languages and unfamiliar cultural habits and religious beliefs frustrate the young physician who is attempting to render what would seem to be routine and logical medical care. This is compounded further by the fact that the patient often doesn’t know the doctor, doesn’t understand the hospital system, and is not able to communicate freely to the physician concerning his body, fears, or desires. The patient may also be influenced by the current wave of populism in which physicians are resented as being from another stratum of society. Thus, lack of adequate communication and possible mutual dislike between patient and physician may flaw the relationship. New house officers have no experience in dealing with the complex sociocultural matrix surrounding both traditional medicine and contemporary medical education.

Medical Education in the 90s

It is not within our province as physicians to solve these enormous social problems. But we must be educated about them. We must acknowledge cultural diversity and unfamiliar attitudes and beliefs. Despite their own individual backgrounds, physicians must be ready to accept “difference” and perceived strangeness (7) and to understand that our system of allopathic medicine is not always the first chosen by our patients.

Undergraduate Education

As contemporary society changes, the undergraduate medical curriculum must change to address the emerging needs of the community. The recent influx to the United States of patients with diverse racial, cultural, and religious backgrounds has given rise to the need for a broad sociocultural component to medical education if students are to deal effectively and sensitively with the cultural issues which influence the patient-physician encounter and which affect even the manifestation of disease itself.

Teaching the interpersonal skills needed for effective clinical practice must take into account the broadening cultural diversity of the student and housestaff population. The Mount Sinai School of Medicine has begun a program entitled

"Cultural Diversity in Medicine," which offers an elective, "Providing Health Care to People of Other Cultures."[1] This course emphasizes the differences in values and beliefs of various cultures within American society and in its immigrant groups. It also addresses how cultures define well being, health, illness, and death, as well as expectations patients and physicians have about how health care is delivered. The course is team-taught by a multicultural, interdisciplinary faculty who address cultural differences in verbal and nonverbal communication styles, as they relate to the patient's interview and examination. They also model the communication skills necessary in multicultural medical settings.

Role playing and interviews with patients and families highlight how individual attitudes, values, and beliefs concerning culture, race, gender, class, and age affect both the physician and the patient in the medical interview. The student comes to recognize how his or her own background influences the interaction. Issues of stereotyping, prejudice, and racism are also addressed. In turn, the student can appreciate how the patients' backgrounds also affect perceptions about the health care system and the processing of medical information. Small groups work best for this kind of program, as we discovered in our approach to teaching humanistic medicine (2). The elective program is expanding and has been introduced into the second-year curriculum in the unit on critical assessment of the patient called "Evaluating Patients of Different Cultural Backgrounds."

Graduate Education

Issues concerning HIV, substance abuse, and mental illness have recently been incorporated into rounds at a variety of levels, but multidisciplinary medical rounds should continue to be the centerpiece of teaching about sociocultural issues in relation to disease. The individual patient provides the best focus for integration of all relevant aspects. For our school the training model was established in the humanistic medicine program for interns and residents (2). An ethicist, a liaison psychiatrist, and an internist working together as an educational team work through problems of different beliefs, attitudes, and understanding about medicine, hospitals, and physicians. If a social scientist (sociologist, anthropologist) can meet regularly with such a group this would be extremely helpful in teaching how the healing arts are understood and practiced in other societies. At our own school, rounds occur regularly on different units. House staff select the patient presenting the medical or other problem most vexing to them, and the group, which comprises both third- and fourth-year medical students and house staff at the first- and third-year levels, thoroughly explore together the issues raised.

Each medical school or hospital will emphasize sociocultural factors according to its own patient population. We must move quickly to incorporate the educational content needed to address these issues, for another "tidal wave" is upon us which will affect young and mature physicians alike, namely, managed care, which will only intensify the problems we describe.

References

1. A complicating factor: doctors' feelings as a factor in medical care. Distributed by Transit Media Communications. D Hollywood Avenue, Hohokus, NJ 07423.
2. Gorlin R, Zucker H. Physicians' reactions to patients: A key to teaching humanistic medicine. N Engl J Med 1983; 308:1059–1063.
3. Morganthau T. America: still a melting pot? Newsweek 1993 Aug 9:16–24.
4. News of the week in review. New York Times 1993 Sep 19.
5. Clymer A. Moynihan on health: cure social ills. New York Times 1993 Oct 20.
6. Goldsmith J. Hospital/physician relationships: A constraint to health reform. Health News 1993 Fall.
7. Romanucci-Ross L, Moerman DE, Tancredi LR, editors. The anthropology of medicine. From culture to method. 2nd ed. New York: Bergin & Garvey, 1991.

---

[1] Directed by Dr. Edward Poliandro and cotaught by a multiracial, multiethnic faculty: C. Burnett, A. Diaz, P. Elbirt, U. Kim, J. Goldfarb, M. Diaz, and J. Soohoo.

Could Education Be Dangerous to Your Health?

Pitfalls of Misapplied Statistics

Laszlo Sarkozi, Ph.D., Ellis Jacobs, Ph.D., Donald A. Smith, M.D., M.P.H., and Arthur Nizza, D.S.W.

Abstract

It is important to evaluate cardiovascular risk factors. A correlation between cholesterol levels and the duration of formal education has not been observed previously. In analyzing our data, it was found that these two variables had a significant correlation ($p < 0.05$). Further evaluation did not substantiate this correlation. After considering various possible reasons we concluded that the seeming correlation was caused by inappropriate statistical evaluation.

In February 1993, we conducted a "cholesterol/lipid profile screening," organized under the Wellness and Health Promotion program for our employees. About 10% of our employees, 954 individuals, participated within a period of 2 days. Fifty-nine questionnaires were not completed properly and had to be excluded from our results. The purpose of the program was to alert individuals with abnormal results and possible cardiovascular risk. These individuals would be encouraged to see a physician and/or obtain nutritional counseling. At some future date, the results of this survey were to be compared with a follow-up survey to study the effectiveness of the Wellness and Health Promotion program. We evaluated all the available information and compared it with other published data. Significant increase in cardiovascular risk factors was observed and was apparently correlated with increased number of years of education. This correlation had not been observed previously.

Materials and Methods

Demographic data and venous blood samples were obtained from each interested employee during the screening program. Our survey questionnaire included 32 variables: age, gender, ethnicity [African American, Asian (Chinese, Filipino, Indian, Japanese, Korean), Hispanic, Middle Eastern, West Indian (non-Hispanic), White (non-Hispanic)], union status (1199, faculty, nurses), levels of education, height, weight, time elapsed since last meal, history of documented elevated cholesterol, medication, heart attack, angina, heart disease, positive stress test, stroke, high blood pressure, diabetes, smoking, diet, and aerobic exercise. Blood samples were collected, centrifuged, and the serum was separated within 2 hours after collection. Total cholesterol, high-density lipoprotein (HDL) cholesterol, and triglyceride measurement were performed on a Technicon Chem 1 system (Bayer Diagnostics, Tarrytown, NY) according to the manufacturer's recommendation.

Results

Mean values for serum total cholesterol, lipoprotein cholesterol, and triglyceride levels for various ages and sexes, for individuals considered high risk, and distribution based on the number

---

From the Departments of Pathology, Medicine and Social Services, The Mount Sinai Medical Center, New York. Address correspondence and reprint requests to Laszlo Sarkozi, Ph.D., Professor of Pathology, Director, Department of Chemistry, Box 1038, The Mount Sinai Medical Center, One Gustave L. Levy Place, New York, NY 10029-6574.

TABLE

Distribution of Risk Factors among Various Groups of Participants

			Serum Cholesterol (mg/dL)				
Population group	*N*	%	Total	HDL	LDL	Ratios Total/HDL	Serum Triglycerides (mg/dL)
All groups	895		193	45	117	4.5	148
Sex/age (yr)							
Females (total)	581	65	191	49	115	4.1	131
20–34	182	31	170	49	99	3.6	111
35–44	152	26	185	48	110	4.0	127
45–54	161	28	200	49	123	4.3	142
55–64	71	12	227	51	143	4.6	157
>65	15	3	232	41	140	5.3	173
Males (total)	314	35	197	39	122	5.4	181
30–34	104	33	184	38	114	5.1	161
35–44	94	30	203	39	129	5.4	177
45–54	74	24	209	37	126	5.0	209
55–64	35	11	196	40	122	5.1	152
>65	7	2	196	45	119	4.6	160
High-risk factors							
Tot. cholest. >240	18	2	317	40	213	8.6	292
HDL <25	18	2	196	22	85	8.8	465
Tot. cholest./HDL >8	33	4	256	27	153	9.5	387
LDL >200	22	3	305	43	231	7.6	164
Triglycerides >400	28	3	228	29	97	5.1	514
BMI >35	21	2	196	41	127	4.9	140
Education, (yr)							
≦12	144	16	201	43	123	5.0	176
16	116	13	186	45	112	4.3	135
≧20	195	22	198	46	122	4.7	152

Tot. cholest., total cholesterol; BMI, body mass index.

of years of education, are shown (Table). Results of our survey are in agreement with most published reports (1–3), but our results of cholesterol/lipid levels relating to years of education necessitated further evaluation.

Discussion

A surprising observation is presented in Figure 1. Cholesterol, triglycerides, and low-density lipoprotein (LDL) cholesterol levels for individuals with 16 years of education (considered equivalent with a bachelor's degree) ($n = 116$) are less than for individuals with more than 20 years of education (doctoral degree) ($n = 196$). This illustration would make the front page of most newspapers.

We performed various statistical tests. The Kolmogorov-Smirnov normality test indicated that the data matched the pattern expected if the data were drawn from population with normal distribution. The result of the Student's *t*-tests between the groups was $p < 0.05$, indicating that the difference in the mean values of the groups is greater than would be expected by chance. These are statistically significant differences. It is the truth, but not the whole truth. It was said that in many situations what you don't do is more important than what you do.

Our error in interpreting Figure 1 is revealed in Figure 2, which illustrates, depending on sam-

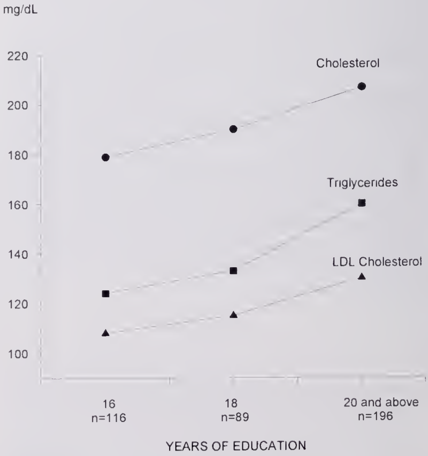


Fig. 1. All lipid risk factors appear to be increasing with increasing number of years of education.

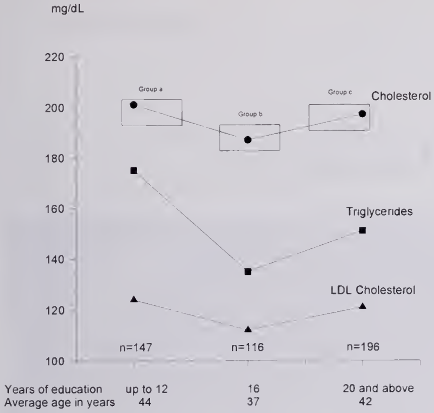


Fig. 2. The trends in cholesterol levels of subpopulation groups of a, b, and c are different than the whole population. The average age of each educational group (44, 37, and 42 years) provides the answer to the misconception of Figure 1.

pling, an inverse relationship (up to 16 years of education, group a), no relationship (14–18 years of education, group b), or a direct relationship (16–20 and more years of education, group c). When we calculated the average age of each educational group, it became obvious that we were presenting age-related cholesterol, triglycerides, and HDL cholesterol and not "education-related" levels. In fact, the group with the least education is at the highest risk because in our survey that group's average age is the highest. Finally, when we grouped the participants by age and plotted their cholesterol levels versus years of education, their average cholesterol levels remained relatively constant, regardless of the number of years of education (Fig. 3).

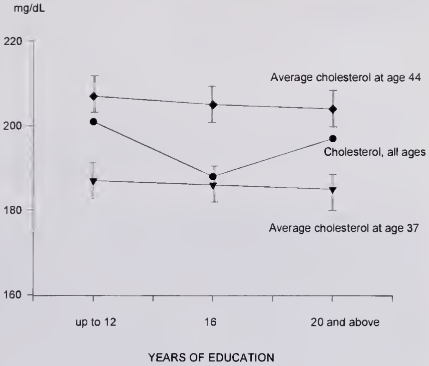


Fig. 3. Cholesterol levels at ages 44 and 37 versus average years of education. The differences are not significant ($p < 0.1$) when comparing years of education; however, the difference between the two age groups is significant ($p < 0.05$).

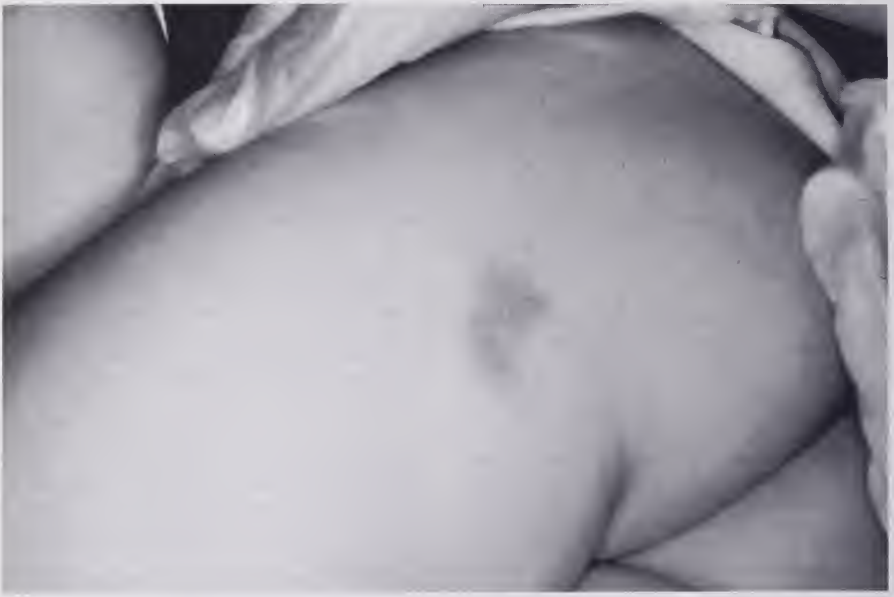
In conclusion, the survey enabled us to review information that has not been available to our Wellness Program. It may be valuable to compare these results with those obtained in the future.

References

1. Johnson CL, Rifkind BM, Sempos CT, et al. Declining serum total cholesterol levels among US adults. JAMA 1993; 269:3002–3008.
2. Sempos CT, Cleeman JI, Carroll MD, et al. Prevalence of high blood cholesterol among US adults. JAMA 1993; 269:3009–3014.
3. Grover SA, Palmer CS, Coupal L. Serum lipid screening to identify high-risk individuals for coronary death. Arch Intern Med 1994; 154:679–684.

Dermatology Notes

Genital Herpes



Genital herpes. Note grouped pustules on an erythematous base. The patient is a young girl who was sexually abused by her father—*Donald Rudikoff, M.D., and Mark Lebwohl, M.D.*

LEVY LIBRARY—MOUNT SINAI

Book Reviews

Effects of Atomic Radiation: A Half-Century of Studies from Hiroshima and Nagasaki

William J. Schull. New York: Wiley-Liss, 1995. 397 pp, $45.00.

THIS BOOK REPRESENTS an important contribution to the literature describing the effects of radiation on survivors of the atomic bomb attacks on Hiroshima and Nagasaki. Although there are numerous articles in both the scientific and general literature reporting these findings, this text represents one of the few successful efforts to describe this work for both the scientist and nonscientist. This was accomplished largely through a clear and lucid writing style as well as limiting the use of scientific terminology so that the nonprofessional can achieve an understanding of this subject. Dr. Schull has gone well beyond the superficial treatment that the scientific aspects of this subject generally receive in the popular media, however. Also, from the perspective of a radiation biologist who has lectured on this subject for many years, this book maintained constant interest because it provided informative and often fascinating explanations about how this research was accomplished and descriptions of the personalities involved in this work, in addition to the science presented. Certainly the author is in an excellent position to write on this subject because he has been an important scientific figure on this work almost since its inception 50 years ago. In addition, the extensive use of notes and references indicated throughout the text and presented in appendices provides all the necessary technical details that may be of particular interest to scientists.

The book begins by briefly describing the circumstances surrounding the atomic bomb attacks on the Japanese cities of Hiroshima and Nagasaki in 1945 at the end of World War II and then quickly moves on to describe the individuals and organizations that have studied the survivors who received significant doses of radiation. Initially, the Allied Powers assumed responsibility for this work; but, since that early period, this research has been performed jointly with Japanese scientists and continues now as the Radiation Effects Research Foundation. An explanation is then provided about the techniques used to estimate the doses received by the irradiated population. This resulted in tentative values in 1965 that were followed by revised and more precise calculations in 1986. The principal effect of this revision was to substantially lower the neutron dose estimates, particularly for the irradiated population in Hiroshima.

The effects on the survivors are then outlined, with a particular emphasis on epidemiologic evidence obtained from this population for the role of radiation in cancer induction. Although there have been numerous studies of various irradiated populations, it has become clear that investigation of the Hiroshima and Nagasaki survivors is most informative because they constitute a large population consisting of individuals of all ages and both sexes who generally received whole-body irradiation. This is in contrast to other groups that generally represented fairly specific populations often with irradiation of a particular body site associated with treatment of a disease. It is from research performed with the Japanese survivors that the very important radiation risk estimate was derived that approximately 5 to 10% of a population exposed to 1 Sv (1 Sievert = 100 rem) will develop a fatal form of cancer. There is little doubt that as most of the Japanese survivors reach the end of their life-span in the next 30 years, the values that will be obtained for excess fatal cancers in this people which can be ascribed to radiation exposure will serve as the most accurate estimates available for susceptibility to radiation carcinogenesis.

The next portion of the text consists of an evaluation of survivors who were exposed to radiation associated with the atomic bomb attacks while developing *in utero*. This group has been observed carefully; data obtained from these studies clearly demonstrate the particular sensitivity of the central nervous system to radiation, because the main effect observed in this prenatally exposed population has been mental retardation. This information has been invaluable in the development of guidelines for pregnant women facing significant radiation exposure associated with medical procedures. Of particular interest will be the incidence of cancer in this population as it ages because this will provide a firm indication as to the sensitivity of the developing fetus to radiation and do much to resolve the long-standing debate over this issue.

Finally, a discussion of genetic effects is outlined which indicates that the induction of hereditary effects among the children of irradiated survivors has not been clearly demonstrated. Although a negative result, these data demonstrate that humans are no more susceptible than the animal model populations that have been used to provide genetic risk estimates and therefore yield an approximate limit to the risk for genetic effects associated with radiation exposure.

Several additional features of this book are quite helpful in addition to the main text, including a summary section, chronology of events, a glossary, and abundant notes and references. Dr. Schull has succeeded admirably in his goal to provide a text that is accessible to nonscientists while maintaining a high level of interest to members of the scientific community.

Barry S. Rosenstein, Ph.D.
Mount Sinai Medical Center

The Cure of Childhood Leukemia: Into the Age of Miracles

John Laszlo. New Brunswick, NJ: Rutgers University Press, 1995. 287 pp, index, $29.95 ISBN 0-8135-2186-6.

THIS IS A BOOK about the "Bench and Bedside" scientists who played a major role in the remarkable advances in childhood acute lymphocytic leukemia (ALL). Only one was a trained pediatrician, all the other physicians were internists. World War II played a role in this tale. Dr. Gordon Zubrod, one of the heroes of this book, served in the armed forces at Goldwater Hospital in New York City working on drug development in malaria. He saw, first hand, what bringing a concentrated group of scientists dedicated to a specific problem could do (in this case the development of atabrine and chloroquine). He used this insight when he became the clinical director of the new Clinical Center of the National Cancer Institute in Bethesda, Maryland. He hired Drs. Emil (Tom) Frei and Emil J Freireich (no relationship to each other). When one of them, namely J, came up with the "radical" idea of transfusing bleeding leukemic children with fresh platelets, the idea was condemned by the most eminent scientists at the National Cancer Institute and elsewhere. Dr. Gordon Zubrod took the responsibility for seeing that these children received fresh platelets. He also saw to it that a double-blind study was carried out which proved unequivocally that platelet transfusion stopped the bleeding in thrombocytopenic patients.

This book is written for the general public as well as for the medical profession. I believe the book could be incorporated into any history of medicine course. Anyone interested in a career in research should know how time, place, and serendipity, in addition to intelligence, play a role in "breakthroughs." The events are described by the scientists themselves. Many of them are colleagues of mine, but I did not know about their early years in medicine.

The most exciting part of the book to me was the interplay between Gordon Zubrod, Tom Frei, and the "controversial" and irrepressible J. Freireich. This book is about people who felt passionately that they could singly and collectively make a difference in the treatment of children with acute leukemia. They didn't use the word cure. It is about the great joy of getting children into remission and then seeing them all grown up years later. When I first became a hematologist, I took care of children as well as adults with leukemia. I clearly remember a 3-year-old girl in whom acute lymphatic leukemia had just been diagnosed and I had to explain to her parents and grandparents, who spoke mainly Yiddish, about acute leukemia and the then current Acute Leukemia Group B Protocol. A social worker interpreted what I said so that they would clearly understand. I believe that this took at least an hour and a half to two hours to do. Happily, the child went into remission. Several months or so later, a new pediatric hematologist came to Mount Sinai and took over the care of all children. Imagine my delight several months ago when, after many years had passed, the same social worker brought me a picture of a beautiful bride. She told me that the family wanted to make sure that I saw this picture of my former patient now as a married woman. Few things can give anyone greater satisfaction than this. It certainly is something that money can't buy.

All adult patients with leukemia quickly learn that the bone marrow examination is the single most important test they will have. They know that they have to achieve a complete remission before a cure is even possible. It has been my custom when this good news occurs to have the intern, the resident, and the nurses who are taking care of the patient to accompany me when I tell the patient this news. It is always a transcendent moment for me, for the patient, for the family, and for the staff as well.

It is passion that distinguishes scientists who make breakthroughs. It is passion which separates good doctors from great ones. It is this passion that we must cultivate and nurture in young physicians as we approach the millennium.

Janet Cuttner, M.D.
Mount Sinai Medical Center

Dermatologic Clinics

Mark Lebwohl and Michael Zanolli, guest editors. Philadelphia: W.B. Saunders Company, vol. 13, no. 4, October 1995.

IT IS NOT OFTEN that the Journal would undertake to review a periodical. However, this particular issue is a worthwhile reading experience on a very important topic, namely psoriasis. The guest editors, Drs. Lebwohl and Zanolli, are uniquely qualified to oversee the effort of the participating contributors who themselves are world-class authorities on their respective subjects. In their preface, Drs. Lebwohl and Zanolli say that the therapeutic regimens are not familiar to primary care physicians and indicate that the public and the profession in general perceive this to be the case. I would certainly agree with that assessment. What strikes me as crucial, however, is the need for the primary physician, who is often the first resource sought out by the patient, to be aware of how the therapeutic modalities in and of themselves have an impact on the general health and welfare of the patient. Certainly an understanding of the pathogenesis of the disease, its genotypic expression, the immunologic ramifications with respect to the associated arthritis, the psychological aspects, drug interactions, and other therapeutic risks are necessary if the primary physician is to serve the patient with this disease in other than a triage capacity. All of this is covered quite well. The material is well referenced; albeit, if criticism is due, it would be that absence of recent references would suggest that the material was a long time in process.

The book covers 20 titles encompassing epidemiology, pathogenesis, immunopathogenesis, immunotherapy, and a host of other therapeutic approaches including photochemotherapy in a snappy and effective style. It is apparent that the available information does not support uniformity in devising a therapeutic approach, but that, after all, is the basis of the sound practice of the art of medicine. No algorithm will suffice in the management of a patient with this illness.

The therapeutic approaches outlined by each of the contributors will speak for themselves. The list of topical therapies is long and continues to grow. The problems encountered by the patient are legion and advice on limiting these obviously must be welcomed. Similarly, the discussions on phototherapy reveal the problems faced by the therapist as well as the patient. Of course, systemic therapy with the wide variety of drugs clearly requires that the physician have a good grasp of their pharmacology and adverse reactions. The section contributed by Dr. Iona Ginsburg on the psychological aspects of psoriasis is well written and enlightening. This section is a must for those in the profession who have contact with any patient with this unhappy disease.

All in all, I am delighted to be able to recommend reading this compilation of papers. I would hope for the patient's sake, however, that future reports will provide a greater degree of optimism with respect to understanding the pathogenesis and therapeutic rational on the part of the physician.

Sherman Kupfer, M.D.
Mount Sinai Medical Center

Introductory Remarks on the Code of Maimonides

Barry S. Coller, M.D.

In a few minutes you will receive your M.D. degree from The Mount Sinai School of Medicine, and in about a year you will receive your license to practice medicine from New York State or another state, but this is the moment that you enter the profession of medicine. You do so by professing your loyalty to a code of behavior embodied in an oath. Long before there was The Mount Sinai School of Medicine, or New York State, or even the M.D. degree, there was an oath for physicians. It is the centerpiece of our contract with society which gives the physician privileges that others do not have, in exchange for the physician's commitment to use all of her or his energy solely for the benefit of the patient, without regard for personal benefit or convenience. At this time in American medicine, when some wish to subvert the physician-patient relationship to control costs, and others wish to exploit it to obtain profits, the oath is more important than ever. Hold to it dearly, each of you, because it is the basis of your patients' trust in you—and the trust of the public in the medical profession is a necessary precondition for the real health care reform yet to come.

In keeping with the tradition established by Dr. James Glenn and carried on by Dr. Arthur Aufses, Jr., I would like to invite all physicians in the audience to please stand with the graduates and reaffirm their commitment to the "Modified Code of Maimonides."

"The Modified Code of Maimonides"

May the love for my art actuate me at all times; may neither avarice, nor miserliness, nor the thirst for glory or a great reputation engage my mind; for the enemies of Truth and Philanthropy could easily deceive me and make me forgetful of my lofty aim in doing good for my brothers and sisters.

May I never forget that the patient is a fellow creature in pain. May I never consider the patient merely a vessel of disease.

When those wiser teach me, let me be humble to learn for the mind of mankind is puny, and the art of healing is vast. May I have the strength, time and opportunity always to verify and correct what I have learned and to increase my understanding. May I always be able to discover today the errors of yesterday, and to obtain a new light tomorrow on what I think I am sure of today.

I have been sanctioned to care for the life and health of mankind. I am about to fulfill my duties; I am ready for my vocation.

****

Editor's Note

The following notes have been extracted from an article by Rosner which appeared in The Mount Sinai Journal of Medicine (1).

The Authorship of the Physician's Prayer

The Physician's Prayer attributed to Moses Maimonides is a lofty and beautiful prayer which first appeared in print in a German periodical in 1783 (2). The anonymous prayer bears the title "Daily prayer of a physician before he visits his patients, from the Hebrew manuscript of a renowned Jewish physician in Egypt from the twelfth century." The first Hebrew version of the prayer was published by Euchel in 1790 (3). The title of Euchel's paper indicated that Marcus Herz was the author of the prayer: It was translated at his request from German into Hebrew.

Murray M. Rosenberg Professor and Chairman of the Samuel Bronfman Department of Medicine, Mount Sinai School of Medicine.

The Physician's Prayer attributed to Maimonides contains moral and ethical standards by which physicians should conduct their professional lives. It is worthy of the lofty ideals and attitude toward the practice of medicine espoused by Maimonides. The healing of the sick, he states, is not only a moral and ethical imperative but is, in fact, a divine commandment. The daily recitation of this prayer serves to remind the physician to constantly embody the highest code on medical philanthropy and professional ethics.

References

1. Rosner F. Moses Maimonides: correcting two misconceptions. Mt Sinai J Med 1992; 62:165–166.
2. Anonymous. Täglisches Gebet eines Arztes bevor er seine Kranken besucht—Aus der hebraischen Handschrift eines berühumten jüdischen Arztes in Egypten aus dem awölften jahrundert. Deuthsches Museum 1783; 1:43–45.
3. Euchel I. Prayer for the physician as he pours out his anxieties before G-d prior to visiting the sick: composed by Sir Hofrat Professor Herz. Ha-Me'assef 1790; 6:242–244.

List of Journal Reviewers in 1996

Each year this journal relies on its reviewers to actively participate in the selection of those manuscripts considered to be worthwhile, timely, interesting, and factual. The journal staff thanks the following reviewers for their time and intellectual effort.

Arthur Aufses, Jr.
Stephen G. Baum
David H. Bechover
Lawrence Berger
Edward J. Bottone
Lewis Burrows
Christopher Cardoza
Mark R. Chassin
Carmel J. Cohen
Barry S. Coller
Charlotte Cunningham-Rundles
Janet Cuttner
Michael J. Droller
Joseph S. Eisenman
Luis R. Espinoza
Solomon Estren
Marianne C. Fahs
Michael J. Fellner
Raul Fleischmajer
Bruce Freundlich
Frederick Friedman
Harriet S. Gilbert
Michael L. Greenberg
Richard S. Haber
Moshe Haimov
Shalom Z. Hirschman
David S. Hodes
Sheldon Jacobson
Julius Jacobson
Barry W. Jaffin
Andrew S. Kaplan
Burton I. Korelitz
Suzanne Kramer
Isadore Kreel
Joel Kreitzer
Phillip Landrigan
Mark G. Lebwohl
Albert M. Lefkovits
Charles S. Lieber
Roy Marcus
Robert Matz
Charles K. McSherry
Burt R. Meyers
Paulo A. Paciucci
Robert G. Phelps
Robert A. Phillips
Jacob H. Rand
Elliot J. Rayfield
Allan P. Reed
Rosamond Rhodes
Martin H. Savitz
Harry Schanzer
Clyde B. Schechter
Michael J. Serby
Harry Spiera
Rabbi Charles Spirn
Barry D. Stimmel
Nelson N. Stone
Rabbi Moses Tendler
Samuel Waxman

Author, Subject, Title Index to Volume 63

Acquired immune deficiency syndrome, histoplasmosis in: 63(2):136–140
Aledort LM: Hemophilia, yesterday, today, and tomorrow, 63(3&4):225–235
Alport's Syndrome, biopsy in: 63(3&4):178–190
Altchek DD: History of the department of dermatology of the Mount Sinai Hospital, 63(3&4):263–272
Alumni Theme Issue; PART A, 63(3&4):143–281; PART B, 63(5&6):335–369
Amyloidoses, Buxbaum J, 63(1):16–23
Amyloid, senile: 63(1):16–23
Amyloid, systemic: 63(1):16–23
Amyloidoses: 63(1):16–23
An imaginary conversation with the gang of three: A ghostly interview with Burrill B. Crohn, Leon Ginzburg, and Gordon Oppenheimer. by Janowitz HD, 63(1):61–65
Analyzing Survival Data from Clinical Trials and Observational Studies by E Marubini and M Valscecchi. S Wallenstein, reviewer, 63(1):68
Androgen receptor mutations in prostate cancer, Barrack ER, 63(5&6):403–412
Art and science of consultation, cardiology: 63(1):39–43
Art and science of consultation, sleep disorders: 63(1):24–38
Art and science of the psychopharmacotherapy of African Americans, Lawson WB, 63(5&6):301–305
Asch E, reviewer: *The PDR Family Guide to Women's Health and Prescription Drugs:* 63(1):66
Atrial fibrillation, in hyperthyroidism: 63(2):118–125
Autonomy and violence, Bar On B, 63(2):82–86

Bar On B: Autonomy and violence: 63(2):82–86
Bar On B: 63(2):107–110
Barrack ER: Androgen receptor mutations in prostate cancer, 63(5&6):403–412
Baumrin B: 63(2):107–110
Bauer JJ: see Raza J
Benign Prostatic Hyperplasia: 63(1):1–9
Berk PD: see Schluger LK
Biopsy, renal: 63(3&4):178–190
Breast cancer chemotherapy after fifty years: more cures and less mutilation, Greenspan EM, 63(3&4):250–255
Book reviews: 63(1):66–69
    Analyzing Survival Data from Clinical Trials and Observational Studies by E Marubini and M Valscecchi.
    Eccentric and Bizarre Behaviors by LR Franzini, and JM Grossberg.
    How to Quit Smoking Without Gaining Weight by M. Katahn.
    Living Heart Brand Name Shopper's Guide by ME DeBakey et al.
    PDR Family Guide to Women's Health and Prescription Drugs 63(5&6):429–431
    The Cure of Childhood Leukemia: Into the Age of Miracles by John Laszlo.
    Dermatologic Clinics, edited by Mark Lebwohl and Michael Zanolli.
    Effects of Atomic Radiation: A Half Century of Studies from Hiroshima and Nagasaki by William J. Schull.
Brown LK: Consultation in sleep medicine: 63(1):24–38
Buxbaum J: The amyloidoses: 63(1):16–23

Calcification of elastic fibers in Pseudoxanthoma elasticum, Truter S, Rosenbaum-Fiedler J, Sapadin A, Lebwohl M, 63(3&4):210–215
Cardiac disease, antithrombotic therapy, risks, 63(5&6):342–358
Canizares O: 63(3&4):263–272
Cancer, prostate: 63(5&6):403–412
Cancer, thyroid: 63(1):10–15
Thomas C. Chalmers—in memoriam, Holland JF, 63(5&6):370–371
Chargin L: 63(3&4):263–272
Chemotherapy, breast cancer, 63(3&4):250–255
Childhood immunizations 1996, Hodes DS, 63(5&6):413–419
Chu P: see Ruiz S
Churg J: see Reisman L
Circumcision, female genitalia: 63(2):87–89
Clinical importance of genetic polymorphism of drug oxidation, Edeki T, 63(5&6):291–300
Collazo Y, Tam R, Sramek J, Herrera J: Neuroleptic dosing in Hispanic and Asian inpatients with schizophrenia, 63(5&6): 310–313
Coller BS: see Code of Maimonides, 63(5&6):432–433
Colliver JA: see Swartz MH
Commencement address, Mount Sinai School of Medicine, 63(2):112–117
Commentary: Code of Maimonides, Coller BS with notes by F Rosner, 63(5&6):432–433
Congress and the international medical graduate: the need for equity, Stimmel BD, 63(5&6):359–363
Cooper JM, Silberzweig J, Mitty HA: Vena cava filters: available devices and current practises, 63(3&4):273–281
Could education be dangerous to your health? pitfalls of misapplied statistics, Sarkozi L, Jacobs E, Smith DA, Nizza A, 63(5&6):425–427
Crohn BB: 63(1):61–65
Crohn's Disease: 63(1):61–65; 63(3&4):191–201
Cross-cultural Issues in Psychopharmacology, 63(5&6):283–334
Cross-cultural perspectives on violence: is violence always violence? Eiser AR, 63(2):87–89
The Cure of Childhood Leukemia: Into the Age of Miracles by John Laszlo. J Cuttner, reviewer, 63(5&6):430
Cuttner J, reviewer: *The Cure of Childhood Leukemia: Into the Age of Miracles* by John Laszlo, 63(5&6):430
Cyclosporine: 63(3&4):191–201

Dean's Seminar
    G Proteins: Out of the cytoskeletal closet, Rodbell M, 63(5&6):381–386
DeBakey, ME: 63(1):68
Deep vein thrombosis: 63(3&4):273–281
Dengue: 63(3&4):159–166
DePalo LR: see Teirstein AS
Dermatologic Clinics, edited by Mark Lebwohl and Michael Zanolli. S Kupfer, reviewer, 63(5&6):431

Dermatology Notes
  Gordon M, Lebwohl M: vitiligo, 63(2):134
  Rudikoff D, Lebwohl M: genital herpes, 63(5&6):428
Dikman S: see Reisman L

Eastham JH, Lacrao JP, Jeste DV: Ethnicity and movement disorders, 63(5&6):314–319
Eccentric and Bizarre Behaviors by LR Franzini, and JM Grossberg. JJ Strain, reviewer: 63(1):67
Edeki T: Clinical importance of genetic polymorphism of drug oxidation, 63(5&6):291–300
Editorial, Time-warp commentary; back to the future. S Kupfer: 63(1):60
Education, dangerous to health? 63(5&6):425–427
Effects of Atomic Radiation: A Half Century of Studies from Hiroshima and Nagasaki by William J. Schull. BS Rosenstein reviewer, 63(5&6):429
Eiser AR: Cross-cultural perspectives on violence: is violence always violence?: 63(2):87–89
Eiser AR: 63(2):107–110
Emergence of "emerging diseases": a lesson in holistic epidemiology Kilbourne ED, 63(3&4):159–166
Emergency department, violence in: 63(2):105–106
Epilogue, violence and medicine, R Rhodes, 63(2):111
Ethics: Violence and Medicine. 63(2):71–111
Ethnicity and antidepressant response, Sramek J, Pi E, 63(5&6):320–325
Ethnicity and movement disorder, Eastham JH, Lacrao JP, Jeste DV, 63(5&6):314–319
Ethnicity and pharmacogenetics, MW Smith, RP Mendoza, 63(5&6):285–290
Ethnicity and psychopharmacology in Latin America, Ramirez LF, 63(5&6):330–331
Evolution of cardiologic consultation. Kuhn LA: 63(1):39–43
Evolution, microbial: 63(3&4):159–166

Fiel IM: see Schluger LK
Fifty years of experience with antithrombotic therapy in cardiac disease: a 1996 approach based on pathogenesis and risk, Freudenberger RS, Fuster V, 63(5&6):342–358
Finasteride: 63(1):1–9
Fleischmajer R: 63(3&4):263–272
Frank AL: Prevention into the 21st century, 63(3&4):236–240
Franzini LR: see Strain JJ
Freudenberger RS, Fuster V: Fifty years of experience with antithrombotic therapy in cardiac disease: 1996 approach based on pathogenesis and risk, 63(5&6):342–358
Friedman LS, Lynn RB: Irritable bowel syndrome, 63(2):126–133
Fuster V: see Freudenberger RS
From Bench to Bedside and Back
  Perturbing gene expression with oligodeoxynucleotides: research and potential therapeutic applications, Gewirtz AM, 63(5&6):372–380

G Proteins: Out of the cytoskeletal closet, Rodbell M, 63(5&6):381–386
Gender differences in psychotropic medications, Lewis-Hall F, 63(5&6):326–329
Genital herpes, Rudikoff D, Lebwohl M, 63(5&6):428
Gert B: The nature and limits of Violence: 63(2):72–76
Gert B: 63(2):107–110
Gewirtz AM: Perturbing gene expression with oligodeoxynucleotides: research and potential therapeutic applications, 63(5&6):372–380
Ginzberg L: 63(1):61–65
Givelber D: Legal obligation to prevent harm; the Tarasoff Rule and its limits: 63(2):77–81
Givelber D: 63(2):107–110
Glick A: 63(3&4):263–272
Glomerulonephritis, various forms and biopsy in: 63(3&4):178–190
Goldenberg H: 63(3&4):263–272
Gordon M: see Dermatology notes
Gorlin R, Strain J, Rhodes R: Physician's reaction to patients: what has happened during the past 10 years, 63(5&6):420–424
Grand Rounds
  Amyloidoses, 63(1):16–23
  Benign prostatic hyperplasia, 63(1):1–9
  Detection of thyroid cancer, 63(1):10–15
  Heart and thyroid disease, 63(2):118–125
  Peptic ulcer disease: paradigms lost, 63(5&6):387–398
Greenspan EM: Breast cancer chemotherapy after fifty years: more cures and less mutilation, 63(3&4):250–255
Grimm KT: 63(2):107–110
Grossberg JM: see Strain JJ
Guarrera JV: see Schluger LK

Haber RS: Thyroid nodules and the detection of thyroid cancer. 63(1):10–15
Harris MT: see Raza J
Hematuria, biopsy in: 63(3&4):178–190
Hemodynamic changes in hypothyroidism: 63(2):118–125
Hemophilia, Aledort LM, 63(3&4):225–235
Hepatoxicity: 63(5&6):364–369
Herpes, genital: 63(5&6):428
Herrera J: see Collazo Y
Herrera J: see Ruiz S
Highman WJ: 63(3&4):263–272
Histoplasmosis, gastrointestinal: 63(2):136–140
History of the department of dermatology of the Mount Sinai Hospital, Altchek DD, 63(3&4):263–272
Hodes DS: Childhood immunizations 1996, 63(5&6):413–419
Holland JF: Thomas C. Chalmers: In Memoriam, 63(5&6):370–371
How to Quit Smoking Without Gaining Weight by M. Katahn. EN Schachter, reviewer: 63(1):66
Hypertension: a personal odyssey, Mendlowitz M, 63(3&4):256–262
Hyperthyroidism, atrial fibrillation, heart failure in, hemodynamic changes, 63(2):118–125
Hypothyroidism: 63(2):118–125
Hypothyroidism, hemodynamic changes in, ischemic heart disease, molecular basis of myocardial dysfunction, myxedema heart, pericardial effusions in: 63(2):118–125

IgA nephropathy, biopsy in: 63(3&4):178–190
Immune Deficiency Syndrome, see Raza J
Immunosuppressive drugs in the treatment of inflammatory bowel disease, history of: origins at The Mount Sinai Hospital, Korelitz BI, Present DH, 63(3&4):191–201
Immunosuppressive drugs, in inflammatory bowel disease, mortality: toxicity, 63(3&4):191–201
Immunizations: 63(5&6):413–420
Infants, violence against: 63(2):101–102
Inflammatory bowel disease: 63(3&4):191–201
Inflammatory bowel disease, mucosal immune system, pathogenesis: 63(3&4):202–209
Involuntary confinement for tuberculosis control: the Jewish view, F Rosner: 63(1):44–48
Irritable bowel syndrome: Friedman LS, Lynn RB, 63(2):126–133
Ischemic heart disease in hypothyroidism: 63(2):118–125
Isoniazid hepatotoxicity after orthotopic liver transplantation, Schluger LK, Sheiner PA, Jonas M, Guarrera JV, Fiel IM, Meyers B, Berk PD, 63(5&6):364–369
Issues in Medical Ethics: Violence and Medicine: 63(2):71–111

Jacobs E: see Sarkozi L
Jacobson S: Violence in the emergency department, 63(2):105–110
Janowitz HD: An imaginary conversation with the gang of

three: a ghostly interview with Burrill B. Crohn, Leon Ginzburg, and Gordon Oppenheimer: 63(1):61–65
Jeste DV: see Eastham J
Jonas M: see Schluger LK

Katahn M: see Schachter EN
Khan A: Myasthenia gravis, 63(2):135
Kilbourne ED: The emergence of "Emerging Diseases": a lesson in holistic epidemiology, 63(3&4):159–166
Kimray-Greenfield filter: 63(3&4):273–281
Kohlberg's thesis: 63(2):87–89
Korelitz BI, Present DH: A history of immunosuppressive drugs in the treatment of inflammatory bowel disease: origins at The Mount Sinai Hospital, 63(3&4):191–201
Kuhn LA: Evolution of cardiologic consultation, 63(1):39–43
Kupelnick B: see Sacks HS
Kupfer S: reviewer: *Dermatologic Clinics*, edited by Mark Lebwohl and Michael Zanolli: 63(5&6):431
Kupfer S: Editorial, Time-warp commentary; back to the future: 63(1):60
Kupfer S: see Reisman L

Lacrao JP: see Eastham JH
Ladenson PW: The Heart and thyroid disease, 63(2):118–125
Lawson WB: The art and science of the psychopharmacotherapy of African Americans, 63(5&6):301–305
Lebwohl M: 63(3&4):263–272
Lebwohl M: see Dermatology notes
Lebwohl M: see Truter S
Lefkovitz AM: Commentary, Alumni Theme Issue, 63(3&4): 143
Legal obligation to prevent harm; the Tarasoff Rule and its limits, D. Givelber: 63(2):77–81
Letter to the Editor and Author's Reply, Simon T, Perez N. Prostate Cancer Screening, 63(2):141
Lewis-Hall F: Gender differences in psychotropic medications, 63(5&6):326–329
Lifton RJ: The psychology of violence: 63(2):90–96
Lifton RJ: 63(2):107–110
Lin A: see Mayer L
Lin K-M: Psychopharmacology in cross-cultural psychiatry, 63(5&6):283–284
Living Heart Brand Name Shopper's Guide by ME DeBakey et al. L Silver, reviewer: 63(1):68
Lustgarten S: 63(3&4):263–272
Lyme disease: 63(3&4):159–166

M protein, group A streptococci: 63(3&4):144–158
Maimonides, Code of: 63(5&6):432–433
Mayer L, Yio XY, Lin A, Panja A: Pathogenesis of inflammatory bowel disease, 63(3&4):202–209
Martinez RA: Minorities in research, 63(5&6):332–334
Marubini E: 63(1)68
Mendlowitz M: Hypertension: A personal odyssey, 63(3&4): 256–262
Meta-analysis, an update: Sacks HS, Pagano D, Kupelnick B, 63(3&4):216–224
Meyers B: see Schluger LK
Microbes, interaction with host: 63(3&4):159–166
Minorities in research, Martinez RA, 63(5&6):332–334
Mitral valve, dysfunction in hyperthyroidism: 63(2):118–125
Mitty HA: see Cooper JM
Mobbin-Uddin umbrella: 63(3&4):273–281
Morchand center for clinical competence: 63(3&4):241–249
Moros DA, Rhodes R, Robichaud A: Violence and medicine, introduction: 63(2):71
Moros DA: Violence and the physician's duty: panel discussion, 63(2):107–110

Nature and limits of Violence, Gert B, 63(2):72–76
Nature of rheumatogenic streptococci, Stollerman GH, 63(3&4):144–158
Narcolepsy: 63(1):24–38
Need for more primary care physicians to contain costs and improve access: myths versus realities. Stimmel B: 63(1):49–59
Nephrotic syndrome, biopsy in: 63(3&4):178–190
Neuroleptic dosing in Asian and Hispanic outpatients with schizophrenia, Ruiz S, Chu P, Sramek J, Rotavu E, Herrera J, 63(5&6):306–309
Neuroleptic dosing in Hispanic and Asian inpatients with schizophrenia, Collazo Y, Tam R, Sramek J, Herrera J, 63(5&6):301–313
Nizza A: see Sarkozi L
Nodules, thyroid: 63(1):10–15
Novello AC: Mount Sinai Medical School commencement address: May 19, 1995, 63(2):112–117

Oligodeoxynucleotides: 63(5&6):372–380
Ophthalmology Notes
  Myasthenia gravis, A Khan, 63(2):135
Oppenheimer G: 63(1):61–65
Osteoclast and osteoporosis, Teitelbaum SL, 63(5&6):399–402
Oxidation, drugs: 63(5&6):291–300

Padilla ML: see Teirstein AS
Pagano D: see Sacks HS
Paris BEC: Violence against elderly people, 63(2):97–100
Paris BEC: 63(2):107–110
Pathogenesis of inflammatory bowel disease, Mayer L, Yio XY, Lin A, Panja A, 63(3&4):202–209
Panja A: see Mayer L
Pathogenesis, and medical treatment with finasteride, Sherwood L: 63(1):1–9
PDR Family Guide to Women's Health and Prescription Drugs E Asch, reviewer, 63(1):66
Peck S: 63(3&4):263–272
Peptic ulcer disease: paradigms lost, Sarner A, Babyatsky M, 63(5&6):387–398
Perez N: see letter to editor
Pericardial effusion, in hypothyroidism: 63(2):118–125
Perturbing gene expression with oligodeoxynucleotides: research and potential therapeutic applications, Gewirtz AM, 63(5&6):372–380
Pharmacogenetics, ethnicity 63(5&6):285–290
Physician's reactions to patients: what has happened during the past 10 years, Gorlin R, Strain J, Rhodes R, 63(5&6):420–424
Pi E: see Sramek J
Present DH: see Korelitz BI
Prevention into the 21st century, Frank AL, 63(3&4):236–240
Polymorphism, genetic: 63(5&6):291–300
Primary care, affecting cost of medical care: 63(1):49–59
Primary care, diminishing interest of US graduates in: 63(1):49–59
Primary care, need for greater number of primary care physicians: 63(1):49–59
Primary care, role of medical schools: 63(1):49–59
Primary care, vacancies in residency positions: 63(1):49–59
Prostate, cancer: 63(5&6):403–412
Prostate, hyperplasia, benign, endocrine pathogenesis, medical treatment, 63(1):1–9
Prostate screening: see letter to editor
Pseudoxanthoma elasticum: 63(3&4):210–215
Psychodynamic psychotherapy, research in, Wallerstein RS, 63(3&4):167–177
Psychology of violence, Lifton RJ, 63(2):90–96
Psychopharmacology in cross-cultural psychiatry, Lin K-M, 63(5&6):283–284
Pulmonary embolism: 63(3&4):273–281

Rabin E: 63(2):107–110
Ramirez LF: Ethnicity and psychopharmacology in Latin America, 63(5&6):330–331

Raza J, Harris MT, Bauer JJ: Gastrointestinal histoplasmosis in a patient with acquired immune deficiency syndrome, 63(2):136–140
Receptor mutations, androgen: 63(5&6):403–412
Reitman D: see Sacks HS
Renal biopsy: why and when, Reisman L, Dikman S, Churg J, Kupfer S, 63(3&4):178–190
Reisman L, Dikman S, Churg J, Kupfer S: Renal biopsy: Why and when. 63(3&4):178–190
REM, intrusive phenomena: 63(1):24–38
Rheumatic fever: 63(3&4):144–158
Rhodes R: see Gorlin R
Rhodes R: see 63(2):71
Rhodes R: 63(2):107–110
Rhodes R: Violence and medicine, epilogue, 63(2):111
Robichaud A: see 63(2):71
Robichaud A: 63(2):107–110
Rodbell M: G Proteins: Out of the cytoskeletal closet, 63(5&6):381–386
Rosen I: 63(3&4):263–272
Rosenbaum-Fiedler J: see Truter S
Rosenstein BS, reviewer: *Effects of Atomic Radiation: A Half Century of Studies from Hiroshima and Nagasaki* by William J. Schull: 63(5&6):429
Rosner F: see Coller BS
Rosner F: Involuntary confinement for tuberculosis control: the Jewish view: 63(1):44–48
Rotavu E: see Ruiz S
Ruiz S, Chu P, Sramek J, Rotavu E, Herrera J: Neuroleptic dosing in Asian and Hispanic outpatients with schizophrenia, 63(5&6):306–309

Sacks HS, Reitman D, Pagano D, Kupelnick B: Meta-analysis: an update, 63(3&4):216–224
Sapadin A: see Truter S
Sarcoidosis mythology, Teirstein AS, Padilla ML, DePalo LR, Schilero GJ, 63(5&6):335–341
Sarkozi L, Jacobs E, Smith DA, Nizza A: Could education be dangerous to your health? Pitfalls of misapplied statistics, 63(5&6):425–427
Sarner A, Babyatsky M: Peptic ulcer disease: paradigms lost, 63(5&6):387–398
Schachter EN, reviewer: *How to Quit Smoking Without Gaining Weight* by M. Katahn, 63(1):66
Schilero GJ: see Teirstein AS
Schluger LK, Sheiner PA, Jonas M, Guarrera JV, Fiel IM, Meyers B, Berk P: Isoniazid hepatotoxicity after orthotopic liver transplantation, 63(5&6):364–369
Seremetis, SV: Violence and women: case presentation and discussion, 63(2):103–104
Seremitis SV: 63(2):107–110
Shatin H: 63(3&4):263–272
Sheiner PA: see Schluger LK
Sherwood LM: Benign prostatic hyperplasia: natural history, endocrine pathogenesis, and medical treatment with finasteride: 63(1):1–9
Silberzweig J: see Cooper JM
Silver L, reviewer: *The Living Heart Brand Name Shopper's Guide* by ME DeBakey et al. 63(1):68
Simon T: see letter to editor
Sleep disorders, associated with medical/psychiatric disorders: 63(1):24–38
Sleep disorders, management: 63(1):24–38
Sleep, dyssomias: 63(1):24–38
Sleep, parasommias: 63(1):24–38
Smith DA: see Sarkozi L
Smith MW, Mendoza RP. Ethnicity and pharmacogenetics, 63(5&6):285–290
Sramek J: see Collazo Y
Sramek JJ, Pi EH: Ethnicity and antidepressant response, 63(5&6):320–325
Sramek J: see Ruiz S
Standardized patients: 63(3&4):241–249
Statistics, misapplied: 63(5&6):425–427
Stimmel BD: The need for more primary care physicians to contain costs and improve access: myths versus realities: 63(1):49–59
Stimmel BD: Congress and the international medical graduate: the need for equity, 63(5&6):359–363
Stollerman GH: The nature of rheumatogenic streptococci, 63(3&4):144–158
Strain J: see Gorlin R
Strain JJ, reviewer: *Eccentric and Bizarre Behaviors* by LR Franzini, JM Grossberg: 63(1):67
Streptococci, rheumatogenic: 63(3&4):144–158
Swartz MH, Colliver JA: Using standardized patients for assessing clinical performance: an overview, 63(3&4):241–249
Suffering, listening to accounts of: 63(2):82–86

Tam R: see Collazo Y
Tarasoff decision: 63(2):77–81
Tarasoff rule, implications: 63(2):77–81
Teets-Grimm K: Violence against infants, 63(2):101–102
Teirstein AS, Padilla ML, DePaolo LR, Schilero GJ, Sarcoidosis mythology, 63(5&6):335–341
Teitelbaum SL: The osteoclast and osteoporosis, 63(5&6):399–402
Thyroid nodules and the detection of thyroid cancer, RS Haber, 63(1):10–15
Thyroid cancer: 63(1):10–15
Thyroid nodules: 63(1):10–15
Toxic shock syndrome, streptococcal: 63(3&4):144–158
Transplantation, liver: 63(5&6):364–369
Transplantation, renal, biopsy in: 63(3&4):178–190
Transthyretin: 63(1):16–23
Truter S, Rosenbaum-Fiedler J, Sapadin A, Lebwohl M: Calcification of elastin fibers in Pseudoxanthoma Elasticum, 63(3&4):210–215
Tuberculosis control, Jewish view of involuntary confinement: 63(1):44–48
Tubulointerstitial nephritis, biopsy in: 63(3&4):178–190

Ulcerative colitis: 63(3&4):191–201
Using standardized patients for assessing clinical performance, Swartz MH, Colliver JA, 63(3&4):241–249

Valscecchi M: 63(1):68
Vena cava filters: available devices and current practises: Cooper JM, Silberzweig J, Mitty HA, 63(3&4):273–281
Violence against elderly people, 63(2):97–100
Violence, against infants: 63(2):101–102
Violence, in emergency department: 63(2):105–106
Violence, intervention strategies: 63(2):97–100
Violence, limits: 63(2):72–76
Violence and medicine, introduction, 63(2):71
Violence, nature: 63(2):72–76
Violence and the physician's duty: panel discussion, DA Moros 63(2):107–110
Violence, pleasure of: 63(2):82–86
Violence, against women: 63(2):103–104
Violent, right to be: 63(2):82–86
Vitiligo: 63(2):134

Wallenstein S, reviewer: *Analyzing Survival Data from Clinical Trials And Observational Studies* by E Marubini and M Valscecchi: 63(1):68
Wallerstein RS: Research in psychodynamic psychotherapy, 63(3&4):167–177
Women, violence against: 63(2):103–104

Yio, XY: see Mayer L

Information for Authors

THE MOUNT SINAI JOURNAL OF MEDICINE, established in 1934 as *The Journal of the Mount Sinai Hospital,* is a peer-reviewed general medical journal. It is indexed or abstracted in *Index Medicus, Current Contents, Science Citation Index, Hospital Literature Index, International Nursing Index, Excerpta Medica, Chemical Abstracts, Biological Abstracts,* and major databases. Theoretical and basic science articles for the clinician, reviews, reports of clinical or research experience, articles on any subject affecting the practice of medicine, short notes in any specialty, book reviews, letters, and articles by medical students are welcome. The *Journal* is a participant in the Agreement on Uniform Requirements for Manuscripts Submitted to Biomedical Journals. The *Journal* aims at responding to authors of manuscripts within 8 weeks of submission.

Communications. Address correspondence on manuscripts to: Editor, *The Mount Sinai Journal of Medicine,* The Mount Sinai School of Medicine, Box 1094, One East 100th Street, New York, NY 10029-6574. Phone: (212) 241-6108; FAX: (212) 722-6386; e-mail: msjournal@msvax.mssm.edu.

Authorship Responsibility, Financial Disclosure, and Assignment of Copyright. Each author is required to sign (1) the statement on authorship responsibility; (2) the statement on financial disclosure; and (3) either the statement on copyright transfer or the statement indicating that the author is or was a US government federal employee when this work was conducted and prepared for publication. These statements are printed from time to time on a page in the *Journal* entitled "Required Form for Authors" and can be photocopied for use. Copies may also be obtained from the *Journal* office. Use additional copies of the form if there are more than six authors, and replace with the appropriate number of the author, e.g., 7th, 8th, etc. The signed original form(s) should be submitted with the manuscript.

Manuscript Preparation. Submit manuscripts in triplicate, including three sets of illustrations (see below). **Manuscripts must be typewritten or computer generated, double spaced throughout (including references, legends, and tables),** on one side of 8½ × 11 inch white bond paper with generous margins. All pages except the title page should be numbered consecutively. Begin a new page for the title page, abstract, initial page of text, references, each table, and legends. A floppy disk, if available, should accompany the manuscript. The label on the disk should specify the name of the first author, the title of the manuscript, how the disk has been formatted, and the word processing system used (DOS, Windows, or Windows 95). PC format is preferred but a high-density formatted Macdisk is acceptable. The text should be in ASCII, Microsoft Word or Wordperfect. Other than the title page, the abstract page and each page of text should be numbered sequentially, and include a heading (top right) which identifies the last name of the first author.

Drugs must be identified by their generic names. Do not use abbreviations on the title or abstract page. Define all abbreviations on first use in the body of the manuscript.

Any material taken directly from other sources must be accompanied by a written statement from both the author and publisher giving permission to the *Journal* for reproducing it. The authors must document in the manuscript that all clinical investigations described in the manuscript were conducted in accord with the guidelines of the Declaration of Helsinki and have been approved by the Institutional Review Board or its equivalent. The text must indicate that the informed consent of each subject was obtained. If results of experiments on animals are reported in the manuscript, the text must indicate that these studies were conducted in accord with the National Institutes of Health (US) guidelines for the care and use of laboratory animals and have been approved by the institutional office or its equivalent which supervises such work.

Title Page. The title page should be double spaced, and not numbered. It should indicate (1) title of the article, phrased as concisely as possible; (2) last and first name of each author, middle initials, if any, and not more than two academic degree(s); (3) institutional affiliation and title or rank of each author; (4) name and address of department(s) and institution(s) from which the work originated; (5) the name and address (and FAX number if available) of the author who will receive and respond to correspondence; (6) the name and full postal address of the author from whom reprint(s) are to be requested; (7) acknowledgments of sources of grant support; (8) name, place, and date of any conferences at which some or all of the information was presented; and, (9) a running head of fewer than 40 letter spaces. At the bottom of the title page, *list 3 to 10 key words or phrases* to facilitate indexing in accord with Medical Subject Headings from Index Medicus.

Abstract. All articles must include an abstract of not more than 250 words on the first numbered page of the manuscript. Abstracts should be double spaced and divided into four paragraphs:

Background: what problem(s) was(were) addressed;
Methods: how the study was performed;
Results: the salient results;
Conclusions.

Key Words. Three to 10 key words or phrases must be provided at the bottom of the title page to facilitate indexing. (see above)

Units of Measurement. SI units are preferred, but values may be expressed in conventional units and conversion factors indicated. The range of normal values for the reporting laboratory must be indicated. When conventional units are used in figures or tables, the appropriate conversion factor(s) must be provided in legends or footnotes.

Editing. All papers will be edited to enhance conciseness and clarity without altering meaning and to ensure conformity with journal style. Authors whose primary language is not English should obtain help from someone proficient in English spelling and grammar.

Illustrations. Photographs, line drawings, and graphs should enhance understanding of the text. Three copies of each must be submitted with the manuscript. Line drawings and graphs should be professionally prepared or computer generated and camera ready. Symbols, lettering, and numbers should be large enough to remain easily legible after the figure has been reduced to fit the width of one column. Photographs must be sharp, high contrast, black and white glossy prints not exceeding 5 inches in height × 7 inches in width. The authors must pay attention to the quality of print reproductions of x-ray plates or transparencies to assure that the special features are labeled and will remain easily visible when reduced to fit the width of one column. The name of the subject and hospital number(s) should not be reproduced in the print. Black and white copies of photomicrographs are preferred and must be cropped to a width of 3⅛ inches. The type of stain and power of magnification must be indicated in the relevant legend. Electron photomicrographs should have internal scale markers. On the back of each print and photograph affix a gum label on which the number of the figure has been written to correspond to its identification in the text, title of manuscript, first author's last name, and the word "top" and arrow to indicate the top edge. **Do not write this information directly on the print and do not mount the print on cardboard.** Illustrations which are to appear together should be mounted accordingly as plates and should be equally sized. Protect the prints by enclosing them in cardboard and separate envelopes. **Do not use paper clips.**

Every effort should be made to avoid using a photograph which includes the face or part of the face of a subject. Facial features should be marked to prevent identification of the subject unless the illustration carries details absolutely essential to the text. **In any case, a partial or full facial photograph of a**

subject will not be published unless written permission has been obtained from the subject and accompanies the photograph.

Legends. Legends should not duplicate the text. Each should be on a separate sheet of paper and double spaced. Number each legend to match the illustration it describes. Illustrations mounted as plates to appear together should be accompanied by a single legend identifying the separate elements by position ("upper left," "middle") or by letters. If such illustrations have been labeled with letters, the same letters are to be used in the legend. The legend must indicate the magnification and type of stain used for photomicrographs.

Tables. Each table should be typed on a separate page, double spaced. Number tables consecutively (Table 1, etc.) to correspond to their appearance in the text. Give each table a brief title that is self-explanatory without reference to the text and type the title on a separate line below its number. Identify all abbreviations and/or symbols used in the table in a footnote.

References. All references cited in the text are to be numbered sequentially. Enclose the citation number in parentheses and place it before any punctuation, e.g., Cytomegalovirus (1), steroids (2,3), and recreational drugs (4–6) have all been implicated (7). Do not use superscripts. The reference list must be limited to works cited in the text, and numbered in the order in which they are cited. Type, double spaced on separate pages following the arrangement and punctuation style in the examples below. Journal title abbreviations must conform to those used in *Index Medicus*.

1. Brown LK. Sleep apnea syndromes: overview and diagnostic approach. Mt Sinai J Med 1994; 61:99–112.

2. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. N Engl J Med 1991; 324:424–428.

3. Couser WG. Glomerular diseases. In: Stein J. ed. Internal Medicine 2nd ed. Little Brown: Boston, 1987: 834–861.

List the name of all authors when there are four or fewer. When there are more than four, list the names of the first three, then *et al.*

References to personal communications and unpublished data should not be included in the list of numbered references. Instead, such information should be placed in the text in parentheses. Permission must be obtained from the appropriate individual to so quote the material. This permission must accompany the manuscript at the time of its submission. When the manuscript is accepted for publication, the author(s) will be asked to confirm the accuracy of all references.

Proofs. Proofs are sent to the corresponding author from the printer. Stylistic changes should not be made. Because of increased printing charges, the cost of author's alterations beyond routine correction of typographical errors in text and data will be billed to the author. Proofs should be corrected and returned to and received at the editorial office within 48 hours by express delivery at the author's expense to the address listed above ("Communications").

Reprint Orders. A reprint order form is sent to the corresponding author with proofs; if reprints are desired, the form must be returned to the editorial office, preferably with the proofs, but not later than 3 days to avoid late charges. A check for the exact amount in US dollars should be made out to the order of The Mount Sinai Journal of Medicine. Please indicate on the check the last name of the first author, and the title of the article. If for one reason or another, the reprint order form is not received with the proofs, it may be obtained directly from the editorial office via telephone or FAX. Reprints are available in minimum quantities of 100, at a cost depending on the number of printed pages and whether covers are desired. This information is directly available from the editorial office.

Required Form for Authors:

AUTHORSHIP RESPONSIBILITY, FINANCIAL DISCLOSURE, AND ASSIGNMENT OF COPYRIGHT

Article Title (type):

Each author must read and sign this form dealing with (1) authorship responsibility, (2) financial disclosure and (3) copyright transfer. An author who was a US federal employee when this work was conducted and prepared for publication, must sign statement (4) instead of statement (3).

1. Authorship Responsibility

I certify that I have participated sufficiently in the conception and design of this work, and the analysis of the data (where applicable), as well as the writing of the manuscript, to take public responsibility for it. I believe that the manuscript represents valid work. I have reviewed the final version of the manuscript and approve it for submission for publication. Neither this manuscript nor one with substantially similar content under my (our) authorship has been published or is being considered for publication elsewhere, except as described in a separate attachment. Furthermore, I agree to produce the original recorded data on which the manuscript is based for examination by the Editor-in-Chief and/or his assignees should such records be requested.

2. Financial Disclosure

I certify that I do not have and have not had an affiliation or financial involvement with or in any organization or entity with a direct financial interest in the subject matter or materials being discussed in the manuscript (eg, employment, consultancies, stock ownership, honoraria), except as disclosed in a separate attachment.

Author(s) Signature(s) **Date**

1st

2nd

3rd

Author(s) Signature(s) **Date**

4th

5th

6th

3. Copyright

In compliance with the Copyright Revision Act of 1976, the signature of each author on this form shall indicate and be evidence of agreement and mutual understanding with *The Mount Sinai Journal of Medicine*, in consideration of the further action of the *Journal* to review and edit the manuscript being submitted, to transfer, assign, and convey all copyright ownership, including any and all rights incidental thereto, exclusively to *The Mount Sinai Journal of Medicine* if this work is published in the *Journal*.

Author(s) Signature(s) **Date**

1st

2nd

3rd

Author(s) Signature(s) **Date**

4th

5th

6th

4. US Federal Employees

I was an employee of the US federal government when this work was conducted and prepared for publication; therefore, it is not protected by the Copyright Act and copyright ownership cannot be transferred.

Author(s) Signature(s) **Date**

1st

2nd

3rd

Author(s) Signature(s) **Date**

4th

5th

6th

Return the original signed form to The Editor, The Mount Sinai Journal of Medicine, Mount Sinai School of Medicine, Box 1094, One East 100th Street, New York, N.Y. 10029. Retain one copy of the form for your file. You may make as many photocopies as needed if your manuscript has more than six authors.

THE
MOUNT SIN
OF MEDICINE

LEVY LIBRARY, MT. SINAI MEDICAL CENTER
3 4805 0452774 9

Forthcoming:

Grand Rounds

The Supraventricular Tachycardias
M. Noelle S. Langan
The Mount Sinai School of Medicine

Arrhythmias and Antiarrhythmic Drugs
Brian F. Hoffman
College of Physicians and Surgeons
Columbia University

Multisystem Granulomatous Disorders
Geraint James
Royal Free Hospital School of Medicine
University of London

Silicone Breast Implants and
Connective Tissue Disease: An Overview
Harry Spiera and Robert F. Spiera
The Mount Sinai School of Medicine

Theme Issues

Computers in the Doctor's Office
Editors: Ray Cornbill and Joseph Kanny

Laparoscopy
Editors: Barry Salky and Michael Edye

Human Immunodeficiency Virus—AID
Editor: Henry Sacks

Emergency Medicine
Editors: Andrew Jagoda and Lynne Richardson

Religious & Philosophic Perspectives on Access to Health Care
Editors: Rosamond Rhodes, Daniel A. Moros, Allyson Robichaud

Available now:

Maternal and Infant Health: New Perspectives
Kenneth V. Lieberman and
Charles J. Lockwood, editors
volume 61, no. 5, October 1994
88 pages **$15**

Update on Sleep Disorders
Lee K. Brown, editor
volume 61, no. 2, March 1994
96 pages **$15**

Environmental and Occupational Medicine in Clinical Practice
Richard P. Wedeen, editor
volume 62, no. 5, October 1995
96 pages **$15**

[illegible]

MOUNT SINAI
GUSTAVE L.
AND
JANET W.
LEVY
LIBRARY

MOUNT SINAI
GUSTAVE L.
AND
JANET W.
LEVY
LIBRARY

LEVY LIBRARY, MT. SINAI MEDICAL CENTER
3 4805 0456752 1